

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 952 832**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C07H 21/00 (2006.01)

C12N 15/10 (2006.01)

C12P 19/34 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.03.2017 PCT/IB2017/051779**

87 Fecha y número de publicación internacional: **05.10.2017 WO17168332**

96 Fecha de presentación y número de la solicitud europea: **28.03.2017 E 17773408 (4)**

97 Fecha y número de publicación de la concesión europea: **14.06.2023 EP 3436607**

54 Título: **Captura de dianas dúplex enlazadas**

30 Prioridad:

28.03.2016 US 201662313974 P

07.07.2016 US 201662359468 P

18.10.2016 US 201662409633 P

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

06.11.2023

73 Titular/es:

NCAN GENOMICS, INC. (100.0%)

Suite 302 - 2386 East Mall

Vancouver, BC V6T 1Z3, CA

72 Inventor/es:

MARZIALI, ANDREA y

PEL, JOEL

74 Agente/Representante:

DEL VALLE VALIENTE, Sonia

Observaciones:

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

ES 2 952 832 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

Captura de dianas dúplex enlazadas

5 **Campo de la invención**

La invención se refiere generalmente a capturar, amplificar y secuenciar ácidos nucleicos.

10 **Antecedentes**

10 Las plataformas de secuenciación genómica de alto rendimiento generan grandes cantidades de datos a precios asequibles, pero no son suficientemente precisos. Incluso las mejores técnicas de secuenciación tienen tasas de error alrededor del 1 por ciento. Eso se traduce en cientos de miles de errores en la secuencia de un solo genoma humano. La llamada de base inexacta conduce a una desalineación de secuencia y a la identificación de mutaciones. Aunque están disponibles algoritmos de llamada de base y alineación, la calidad se ve afectada negativamente por errores de amplificación y secuenciación.

20 Si bien se han realizado avances en técnicas de amplificación y secuenciación, la llamada de base y la alineación se mantienen llenas de errores. Por ejemplo, en la plataforma de secuenciación anterior, los fragmentos de ADN se unen a un soporte sólido, tal como una pared de canal. Una vez que un fragmento se une al soporte sólido, el fragmento se amplifica y los productos de amplificación se unen al soporte sólido próximo al fragmento de siembra. El proceso se repite hasta que se forma un agrupamiento de productos de amplificación que deben ser idénticos a los fragmentos de siembra. Sin embargo, solo un fragmento de semillas siembra un agrupamiento. Si hay un error en el fragmento de siembra o se comete un error en la amplificación del agrupamiento, el error se repite en toda o parte del agrupamiento. Este error conduce a minimizar una base y complicar la alineación de secuenciación.

25 Para identificar estos tipos de errores, los métodos de secuenciación de código de barras estándar utilizan decenas a cientos de copias de la misma plantilla, o diez a cientos de agrupamientos para crear un conjunto de muestras para comparación. Al aumentar drásticamente el número de copias o agrupamientos, puede determinarse un error. Sin embargo, esta estrategia es costosa y consume ancho de banda de secuenciación.

30 La patente US-2016/122814 A1 se refiere a métodos de secuenciación con fragmentos enlazados. Brent C. Satterfield, "Cooperative Primers", The Journal Of Molecular Diagnostics, US, (20140301), vol. 16, n.º. 2, páginas 163 - 173.

35 **Resumen**

La invención se define por las reivindicaciones. La invención proporciona métodos para aumentar la precisión de la llamada de bases mediante la unión de dos o más fragmentos procedentes de la misma plantilla de partida. Los fragmentos pueden representar las cadenas sentido y antisentido de una molécula de ADN dúplex. Al unir múltiples plantillas, que incluyen, por ejemplo, ambas cadenas de la molécula dúplex, en una sola lectura, se aumenta la densidad de información y se reducen las tasas de error. En realizaciones dúplex, los datos dúplex permiten una diferenciación fácil entre las variantes verdaderas y los errores introducidos en la amplificación o secuenciación (p. ej., errores que una polimerasa podría hacer en un sentido no es probable que se repitan en ambas cadenas mientras que en una variante verdadera sí lo sería). Pueden utilizarse códigos de barras específicos de detección para confirmar la presencia de copias de plantilla sentido y antisentido en un agrupamiento. Las lecturas de secuenciación sentido y antisentido dedicadas pueden utilizarse para diferenciar entre errores introducidos y variantes verdaderas.

50 En ciertas realizaciones, la invención da a conocer métodos de captura de dianas enlazadas para moléculas de ADN dúplex. Se proporcionan métodos de captura diana basados en soluciones así como también métodos de captura de objetivos basados en gotas. Los métodos basados en solución y gotas utilizan sondas de captura de diana enlazadas que incluyen una sonda universal y una sonda específica de diana, en donde las reacciones se producen en condiciones que requieren que la sonda específica de diana se una para permitir la unión de la sonda universal. Debido a que están implicadas múltiples etapas de unión y extensión, la especificidad se mejora sobre la captura de diana de unión única tradicional. A continuación, la sonda universal unida se extiende mediante la utilización de la polimerasa de desplazamiento de cadena para producir copias de las cadenas diana que a continuación pueden amplificarse mediante la utilización de PCR con cebadores universales. Los métodos de la invención reemplazan los flujos de trabajo de PCR-capture-PCR con una sola etapa de PCR y captura. Las sondas de captura enlazadas pueden utilizarse en uno o ambos sentidos del ADN donde se requieran una mayor especificidad e información dúplex. Son posibles múltiples tipos de enlazadores como se describe a continuación. De manera similar a los métodos de captura de dianas basados en soluciones de la invención, los métodos basados en gotas que permiten a un usuario realizar la captura de la diana en gotas, en lugar de restringirse a la PCR multiplexada en gotas.

60 Los métodos de captura pueden combinarse con cebadores unidos como se describe en la presente memoria para crear moléculas dúplex enlazadas a partir de gotas. Las nanopartículas que comprenden sondas de captura diana así como cebadores universales pueden utilizarse para capturar regiones dirigidas de un grupo de moléculas enlazadas en 5', convirtiendo solo las moléculas diana en semillas dúplex para agrupamientos de secuenciación.

65 Los métodos de la presente invención tienen aplicaciones en la preparación de muestras y secuenciación. En los métodos de preparación de muestras, la presente invención permite fragmentos idénticos o fragmentos que

representan una cadena sentido y antisentido de un ácido nucleico para unirse entre sí. Una molécula de enlace une los fragmentos, creando un complejo. El complejo puede incluir adaptadores, cebadores y moléculas de unión, además de los fragmentos idénticos o fragmentos dúplex. Además, en algunas realizaciones, el complejo puede incluir múltiples fragmentos idénticos enlazados entre sí. En muestras que tienen bajo contenido de ADN diana, tales como muestras prenatales, al enlazar múltiples fragmentos juntos, los fragmentos pueden amplificarse y secuenciarse con mayor precisión con una identificación fácil de errores de secuenciación y amplificación.

Los fragmentos unidos pueden crearse mediante la amplificación de un fragmento de ácido nucleico con cebadores de amplificación enlazados. En ciertas realizaciones, los sitios de cebado universales pueden ligarse al fragmento diana para crear una molécula plantilla. Los métodos pueden incluir flujos de trabajo de gotas y sin gotas y producir moléculas enlazadas que representan ambas cadenas a aproximadamente al menos un 50 % de tasa. En los métodos de amplificación de gotas, la molécula plantilla puede añadirse a una gota junto con cebadores de amplificación múltiple y cebadores universales enlazados. Los cebadores pueden ser cebadores de amplificación directos e inversos específicos de genes multiplexados. La gota puede someterse a una emulsión o a una amplificación de PCR digital. Los productos amplificados deben unirse a copias del fragmento de ácido nucleico original o copias enlazadas de las cadenas sentido y antisentido de un fragmento de ácido nucleico original, dependiendo de la aplicación. Dos o más cebadores o fragmentos de ácido nucleico pueden unirse mediante un derivado de polietilenglicol, un oligosacárido, un lípido, un hidrocarburo, un polímero o una proteína. En ciertas realizaciones, cuatro o más cebadores o fragmentos de ácido nucleico biotinilados pueden unirse con una molécula de estreptavidina o una nanopartícula funcionalizada. Los cebadores enlazados de la invención también pueden incluir secuencias identificadoras de agrupamiento únicas para garantizar que todas las lecturas de agrupamiento se originen a partir de la misma molécula plantilla enlazada.

Los métodos de la presente invención mejoran la llamada de la base cuando se incorporan en técnicas de amplificación. En los métodos de amplificación tradicionales, los amplicones se crean a partir de una sola plantilla. Si existe un error en el fragmento, el error se propaga en todos los productos de amplificación. En lugar de utilizar una sola plantilla, se utilizan múltiples plantillas o plantillas idénticas que comprenden las cadenas sentido y antisentido de una molécula de ADN dúplex para crear los productos de amplificación. En el caso de que se produzca un error en una de las cadenas de la plantilla, la utilización de múltiples plantillas, en lugar de una sola plantilla, permite que dicho error se identifique en la etapa de secuenciación. Cuando se utilizan ambas cadenas de un fragmento de ADN dúplex como plantillas, los errores pueden diferenciarse de las variantes verdaderas que deben encontrarse en ambas cadenas.

En ciertas técnicas de la invención, mediante la siembra con múltiples plantillas, los errores pueden diferenciarse de las variantes verdaderas mediante una disminución de la calidad de secuenciación en una sola lectura en la posición donde las bases no sean las mismas (una variante verdadera estaría presente en todas las lecturas, proporcionando una señal fuerte). En las realizaciones que siembran una cadena con un sentido y antisentido, las variantes verdaderas y los errores pueden identificarse comparando los resultados de una lectura primer sentido con una segunda lectura antisentido para confirmar la presencia de la variante en ambas cadenas de la plantilla.

Los métodos de la invención pueden incluir crear fragmentos de ácido nucleico enlazados a partir de una sola molécula de partida. Al preparar complejos enlazados de siembra a partir de un solo fragmento de ácido nucleico (p. ej., mediante la utilización del método de PCR de emulsión descrito en la presente memoria), se elimina el riesgo de crear complejos híbridos de dos fragmentos de ácido nucleico diferentes.

Otros métodos están relacionados con la reducción de la interferencia entre agrupamientos en secuencias de alta densidad. Los métodos pueden incluir ligar dos o más adaptadores diferentes con diferentes secuencias de cebadores para permitir la diferenciación de agrupamientos mediante la utilización de diferentes cebadores de secuenciación correspondientes a las diferentes secuencias de cebadores del adaptador.

Los métodos de la invención incluyen estrategias de identificación dúplex para moléculas dúplex enlazadas por gotas. Como se señaló, los métodos de la invención basados en gotas pueden dar como resultado al menos un 50 % de tasa de formación de fragmentos dúplex enlazados (moléculas enlazadas que contienen representaciones de cada lado del dúplex de ADN), por lo que la identificación de esos productos se vuelve importante para omitir los datos de los productos que no son dúplex y recoger los beneficios crecientes de precisión de los productos dúplex. Los métodos de identificación dúplex pueden incluir, por ejemplo, un enfoque de PCR de dos etapas que utiliza dos conjuntos de cebadores con diferentes temperaturas de hibridación donde varios ciclos iniciales se realizan a baja temperatura con cebadores de código de barras específicos de genes para amplificar e identificar cada sentido del dúplex, mientras se añade una cola universal para ciclos posteriores. El número de ciclos de código de barras se limita para evitar etiquetar cada sentido del dúplex con múltiples códigos de barras. Los ciclos posteriores pueden realizarse a alta temperatura a través de cebadores universales porque los cebadores de código de barras no pueden unirse en esas condiciones. Los productos dúplex pueden identificarse a continuación por la presencia de sus códigos de barras específicos de detección durante el análisis de secuenciación y distinguirse de los agrupamientos que no sean dúplex. Por lo tanto, puede apreciarse la mayor fidelidad proporcionada por la siembra de agrupamientos dúplex.

En realizaciones que no se basan en gotas, puede utilizarse un único ciclo de amplificación para crear una molécula dúplex enlazada que tenga tanto las cadenas sentido como antisentido del fragmento original. La molécula dúplex enlazada puede cargarse a continuación directamente en una celda de flujo para su secuenciación, evitando así los

sesgos de secuencia o longitud inducidos por la amplificación o (p. ej., en la secuenciación del genoma completo), así como evitar los errores introducidos por la amplificación y las pérdidas de ácido nucleico debidas a una eficiencia de carga deficiente. Por ejemplo, cuando la eficiencia de carga de un secuenciador puede definirse como: (número de lecturas de salida)/(número de moléculas de entrada capaces de formar lecturas), la eficiencia de carga para Illumina MiSeq es <0,1 %, y es similar para otros instrumentos de Illumina. Esto se debe en gran medida a pérdidas fluídicas, ya que más de 600 ul de muestra se cargan en el secuenciador, mientras que solo ~7 ul se retiene dentro de la celda de flujo en la unión, lo que da como resultado grandes pérdidas de material de partida. Los métodos de carga directa que no se basan en gotas descritos en la presente memoria reúnan estas ineficiencias.

Para las realizaciones de carga directa, así como otras aplicaciones donde el rendimiento de la carga de la celda de flujo y el rendimiento de captura de la diana son importantes, puede ser beneficioso combinar la carga de las celdas de flujo con secuenciación dirigida, para minimizar la pérdida. Dicha combinación simplifica adicionalmente el flujo de trabajo eliminando una etapa adicional. Si bien existen métodos para la captura de objetivos en la celda de flujo, sufren de al menos dos lados inferiores. En primer lugar, no son capaces de secuenciar la región que se captura en la celda de flujo. Para fragmentos cortos tales como ADN libre de células, esto puede tener una gran pérdida de señal. En segundo lugar, son incapaces de capturar moléculas dúplex enlazadas, como se describe en la invención, para su secuenciación. Se describe la captura de diana basada en celdas de flujo de moléculas dúplex. La celda de flujo contiene un sentido de oligos con regiones diana, mientras que el otro sentido está capilarizado y no disponible inmediatamente para la unión. Después de que un sentido de moléculas enlazadas es capturado en la celda de flujo, los otros oligos de celdas de flujo se activan para capturar el otro sentido de los fragmentos enlazados (p. ej., mediante la utilización de una digestión de uracilo, digestión enzimática o luz). La plantilla puede a continuación extenderse y la generación de agrupamientos puede continuar tan normal.

Los métodos de la presente invención mejoran la amplificación en un soporte sólido, tal como en la plataforma Illumina (Illumina, Inc. San Diego, CA) o la plataforma Ion Torrent (Thermo Fisher Scientific Inc., Waltham, MA). En la técnica Illumina, utilizando amplificación de puente, se forman agrupamientos de amplicones. Si existe un error en el fragmento, el error se repite en el agrupamiento. Sin embargo, con la presente invención, los fragmentos enlazados se ponen en contacto con el soporte sólido. Los fragmentos, que pueden ser idénticos o representan cada cadena de una molécula de ADN dúplex, siembran el agrupamiento, lo que da como resultado una fracción de los amplicones totales que se derivan de cada uno de los fragmentos. Esta técnica permite determinar fácilmente un error en la etapa de secuenciación y puede ayudar a llamar a las variantes verdaderas y diferenciarlas de errores de secuenciación o amplificación (p. ej., PCR).

Los métodos de la invención mejoran los procesos de amplificación de multiplexación. En algunas realizaciones de la presente invención, los fragmentos enlazados pueden formarse o introducirse en una gota para su posterior amplificación. Si existe un error en algunos de los fragmentos, el error puede determinarse con los datos de secuenciación sin procesar. Se describe en donde los fragmentos enlazados puedan estar enlazados a una microesfera y a continuación, mediante amplificación, los fragmentos siembran la microesfera con amplicones. Al proporcionar la ventaja de formar una pluralidad de amplicones utilizando múltiples copias del mismo fragmento, la presente invención mejora la llamada de base en una variedad de aplicaciones.

Los métodos de la invención pueden incorporarse en múltiples plataformas de secuenciación. Por ejemplo, en secuenciación tradicional por síntesis, cada base se determina secuencialmente. No se determina un error hasta que se utilicen técnicas bioinformáticas para analizar los datos. Sin embargo, la presente invención permite que múltiples fragmentos de ácidos nucleicos se unan entre sí durante las metodologías de secuenciación. Al analizar múltiples fragmentos simultáneamente, el acuerdo entre las bases indica la precisión, mientras que el desacuerdo entre las bases indicaría un error. Con la presente invención, los errores son determinables a partir de los datos de secuenciación sin procesar, sin la aplicación de bioinformática. Esta técnica utiliza menos copias o agrupamientos, aumenta el rendimiento de secuenciación y disminuye los costes.

Breve descripción de las figuras

Las Figuras 1A-1E representan el proceso de formación de los fragmentos enlazados.

La Figura 2 representa las plantillas enlazadas a un soporte sólido.

La Figura 3 representa un ejemplo de fragmentos enlazados.

Las Figuras 4A-4C representan la extensión y la amplificación de los fragmentos enlazados.

Las Figuras 5A-5C representan la determinación de errores en los fragmentos enlazados.

La Figura 6 representa adaptadores y cebadores.

La Figura 7 representa un proceso para la creación de fragmentos de ácido nucleico enlazados mediante PCR de emulsión.

La Figura 8 representa la creación de cuatro fragmentos enlazados en PCR de emulsión y la hibridación celular de flujo de los cuatro fragmentos enlazados a agrupamientos de secuenciación de semillas.

- La Figura 9 representa adaptadores de ligación útiles en la invención.
- 5 La Figura 10 representa cebadores de PCR digitales útiles en la invención.
- La Figura 11 representa un proceso para la creación de fragmentos de ácido nucleico enlazados en una perla mediante PCR de emulsión.
- 10 La Figura 12 representa la secuenciación Ion Torrent de perlas de ácido nucleico enlazadas útiles para la invención.
- La Figura 13 representa un proceso para crear fragmentos de ácido nucleico enlazados con diferentes adaptadores y sitios de cebadores de secuenciación.
- 15 La Figura 14 representa un proceso para los agrupamientos de secuenciación que tienen diferentes sitios de cebadores de secuenciación para reducir la interferencia.
- La Figura 15 ilustra la interferencia entre los agrupamientos de secuenciación.
- 20 La Figura 16 ilustra un método de la invención basado en gotas para crear ácidos nucleicos dúplex enlazados.
- La Figura 17 representa un cebador unido a modo de ejemplo y cebadores específicos de genes directos e inversos y su utilización.
- 25 La Figura 18 representa un cebador unido a modo de ejemplo y cebadores específicos de genes directos e inversos y su utilización.
- La Figura 19 muestra cebadores específicos de genes útiles en la invención.
- 30 La Figura 20 representa un método de secuenciación con métodos de productos mostrados en la Figura 17.
- La Figura 21 representa un método de secuenciación con métodos de productos mostrados en la Figura 18.
- 35 Las Figuras 22A y 22B muestran un método de formación dúplex basado en gotas que no está unido que utiliza un adaptador de enlace.
- Las Figuras 23A y 23B muestran un método de formación dúplex no enlazado de gota que utiliza dos adaptadores de enlace.
- 40 Las Figuras 24A y 24B ilustran etapas de un método de secuenciación de carga directa utilizando moléculas dúplex enlazadas.
- La Figura 25 muestra etapas ejemplares de un método de unión a celdas de flujo.
- 45 La Figura 26 representa un protocolo ilustrativo de preparación de celdas de flujo fuera de línea.
- La Figura 27 ilustra métodos de captura de objetivos basados en celdas de flujo para moléculas dúplex
- 50 Las Figuras 28A-28E representan etapas en un método de captura y secuenciación de objetivos basados en celdas de flujo ejemplares para moléculas dúplex.
- La Figura 30 muestra las tasas de error de secuenciación utilizando los agrupamientos de siembra individual que se alinearon con un amplicón KRAS.
- 55 La Figura 31 representa un agrupamiento sembrado individualmente para métodos de secuenciación utilizados para producir los resultados mostrados en la Figura 30 y un agrupamiento de siembra doble para métodos de secuenciación utilizados para producir los resultados mostrados en la Figura 43.
- 60 La Figura 32 ilustra un método de llamada de base basado en una única lectura de secuenciación y calidad de señal.
- La Figura 33 muestra un método de llamada de base basado en una comparación de una lectura de secuenciación sentido y antisentido.
- 65 La Figura 34 ilustra métodos ilustrativos de captura de dianas de ácidos nucleicos dúplex enlazados.
- La Figura 35 ilustra los métodos de amplificación de ácidos nucleicos diana capturados y enlazados.

La Figura 36 muestra métodos de captura de diana basados en gotas y producción de ácido nucleico dúplex enlazado.

Las Figuras 37A y B muestran etapas de un método de captura de diana basado en gotas de la invención.

La Figura 38 muestra una nanopartícula que tiene cebadores universales y una cadena que comprende una región diana complementaria a una región de captura de la molécula enlazada a capturar.

La Figura 39 ilustra la unión de la región de captura a la región diana.

La Figura 40 muestra la unión de los cebadores universales a sitios de cebadores universales en la molécula enlazada.

La Figura 41 muestra la extensión de cebador universal mediante la polimerasa de desplazamiento de cadena para producir copias enlazadas a nanopartículas de la molécula diana que comprende ambas cadenas de la molécula unida original.

La Figura 42 muestra una nanopartícula de siembra dual que puede utilizarse para sembrar un agrupamiento en un secuenciador de celdas de flujo como se describe en otra parte de la solicitud.

La Figura 43 muestra las tasas de error de secuenciación utilizando los agrupamientos de siembra doble que se alinearon con un amplicón KRAS.

Descripción detallada

La invención se refiere de manera general a métodos para amplificar y secuenciar ácidos nucleicos uniendo dos fragmentos de ácido nucleico. Estos fragmentos pueden ser dos copias idénticas de un solo fragmento o ambas cadenas de un ácido nucleico dúplex. La utilización de dos fragmentos reduce las tasas de error, aumenta la eficiencia en alineación y reduce los costes de secuenciación.

El ácido nucleico de manera general se adquiere de una muestra o un sujeto. Las moléculas diana para el marcaje y/o la detección de acuerdo con los métodos de la invención incluyen, pero no se limitan a, material genético y proteómico, tal como ADN, ADN genómico, ARN, ARN expresado y/o cromosoma(s). Los métodos de la invención son aplicables al ADN de células completas o a porciones de material genético o proteómico obtenidos de una o más células. Los métodos de la invención permiten obtener ADN o ARN a partir de fuentes no celulares, tales como virus. Para un sujeto, la muestra puede obtenerse de cualquier manera clínicamente aceptable, y las plantillas de ácido nucleico se extraen de la muestra mediante métodos conocidos en la técnica. Generalmente, el ácido nucleico puede extraerse de una muestra biológica mediante una variedad de técnicas tales como las descritas por Maniatis, et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., pp. 280-281, 1982).

Las plantillas de ácido nucleico incluyen ácido desoxirribonucleico (ADN) y/o ácido ribonucleico (ARN). Las plantillas de ácido nucleico pueden ser sintéticas o derivadas de fuentes naturales. Los ácidos nucleicos pueden obtenerse de cualquier fuente o muestra, ya sea biológica, ambiental, física o sintética. Las plantillas de ácido nucleico pueden aislarse de una muestra que contenga una variedad de otros componentes, tales como proteínas, lípidos y ácidos nucleicos que no sean plantillas. Las plantillas de ácido nucleico pueden obtenerse de cualquier material celular, obtenerse de un animal, planta, bacteria, hongo o cualquier otro organismo celular. Las muestras para su utilización en la presente invención incluyen virus, partículas virales o preparaciones. El ácido nucleico también puede adquirirse de un microorganismo, tal como una bacteria u hongo, de una muestra, tal como una muestra ambiental.

En la presente invención, el material diana es cualquier ácido nucleico, incluyendo ADN, ARN, ADNc, ANP, LNA y otros que están contenidos dentro de una muestra. Las moléculas de ácido nucleico incluyen ácido desoxirribonucleico (ADN) y/o ácido ribonucleico (ARN). Las moléculas de ácido nucleico pueden ser sintéticas o derivadas de fuentes naturales. Las moléculas de ácido nucleico pueden aislarse de una muestra biológica que contenga una variedad de otros componentes, tales como proteínas, lípidos y ácidos nucleicos que no sean plantillas. Las moléculas plantillas de ácido nucleico pueden obtenerse de cualquier material celular, obtenerse de un animal, planta, bacteria, hongo o cualquier otro organismo celular. Las moléculas de ácido nucleico pueden obtenerse de una sola célula. Las muestras biológicas para su utilización en la presente invención incluyen partículas o preparaciones virales. Las moléculas de ácido nucleico pueden obtenerse directamente de un organismo o de una muestra biológica obtenida de un organismo, p. ej., de sangre, orina, líquido cefalorraquídeo, fluido seminal, saliva, esputo, heces y tejido. Cualquier muestra de tejido o fluido corporal puede utilizarse como fuente para ácido nucleico para su utilización en la invención. Las moléculas de ácido nucleico también pueden aislarse de células cultivadas, tales como un cultivo celular primario o una línea celular. Las células o tejidos de los cuales se obtienen las moléculas de ácido nucleico pueden infectarse con un virus u otro patógeno intracelular. Además, los ácidos nucleicos pueden obtenerse a partir de muestras que no sean celulares ni tisulares, tales como muestras virales o muestras ambientales.

Una muestra también puede ser ARN total extraído de una muestra biológica, una genoteca de ADNc, un virus viral o genómico. En ciertas realizaciones, las moléculas de ácido nucleico están unidas a otras moléculas diana tales como proteínas, enzimas, sustratos, anticuerpos, agentes aglutinantes, perlas, moléculas pequeñas, péptidos o

cualquier otra molécula y sirven como sustituto de la cuantificación y/o detección de la molécula diana. Generalmente, el ácido nucleico puede extraerse de una muestra biológica mediante una variedad de técnicas tales como las descritas por Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor, N.Y. (2001). Las moléculas de ácido nucleico pueden ser monocatenarias, bicatenarias o bicatenarias con regiones monocatenarias (por ejemplo, estructuras de tallo y bucle). Las proteínas o porciones de proteínas (polímeros de aminoácidos) que pueden unirse a restos de unión de alta afinidad, tales como anticuerpos o aptámeros, son moléculas diana para el marcaje de oligonucleótidos, por ejemplo, en gotas.

Las plantillas de ácido nucleico pueden obtenerse directamente de un organismo o de una muestra biológica obtenida de un organismo, p. ej., de sangre, orina, líquido cefalorraquídeo, fluido seminal, saliva, esputo, heces y tejido. En una realización particular, el ácido nucleico se obtiene a partir de plasma congelado fresco (FFP). En una realización particular, el ácido nucleico se obtiene a partir de tejidos fijados en formalina e incluidos en parafina (FFPE). Cualquier muestra de tejido o fluido corporal puede utilizarse como fuente para ácido nucleico para su utilización en la invención. Las plantillas de ácido nucleico también pueden aislarse de células cultivadas, tales como un cultivo celular primario o una línea celular. Las células o tejidos de los cuales se obtienen los ácidos nucleicos plantilla pueden infectarse con un virus u otro patógeno intracelular. Una muestra también puede ser ARN total extraído de una muestra biológica, una genoteca de ADNc, un virus viral o genómico.

Una muestra biológica puede homogeneizarse o fraccionarse en presencia de un detergente o tensioactivo. La concentración del detergente en el amortiguador puede ser de aproximadamente 0,05 % a aproximadamente 10,0 %. La concentración del detergente puede ser hasta una cantidad en la que el detergente permanezca soluble en la solución. En una realización preferida, la concentración del detergente es entre 0,1 % y aproximadamente 2 %. El detergente, particularmente uno suave que no sea desnaturizante, puede actuar para solubilizar la muestra. Los detergentes pueden ser iónicos o no iónicos. Los ejemplos de detergentes no iónicos incluyen el tritón, como la serie Triton X (Triton X-100 t-Oct-C₆H₄-(OCH₂-CH₂)_xOH, x = 9-10, Triton X-100R, Triton X-114 x = 7-8), octilglucósido, polioxietileno(9)dodecil éter, digitonina, IGEPAL CA630 octilfenil polietilenglicol, n-octil-beta-D-glucopiranosido (betaOG), n-dodecil-beta, Tween 20 polietilenglicol sorbitán monolaurato, Tween 80 monooleato de sorbitán de polietilenglicol, polidocanol, n-dodecil beta-D-maltosido (DDM), NP-40 nonilfenil polietilenglicol, C₁₂E₈ (octaetilenglicol n-dodecil monoéter), hexaetilenglicol mono-n-tetradecil éter (C₁₄EO₆), octil-beta-tioglucopiranosido (octil tioglucósido, OTG), Emulgen, y polioxietileno 10 lauril éter (C₁₄E₁₀). Los ejemplos de detergentes iónicos (aniónicos o catiónicos) incluyen desoxicolato, dodecilsulfato de sodio (SDS), N-lauroilsarcosina y cetiltrimetilamonio (CTAB). También puede utilizarse un reactivo de ion híbrido en los esquemas de purificación de la presente invención, como Chaps, ion híbrido 3-14 y 3-[(3-colamidopropil)dimetilamonio]-1-propanosulf-onato. También se contempla que la urea pueda añadirse con o sin otro detergente o tensioactivo.

Las soluciones de lisis u homogeneización pueden contener además otros agentes, tales como agentes reductores. Los ejemplos de tales agentes reductores incluyen ditioneitol (DTT), beta-mercaptoetanol, DTE, GSH, cisteína, cisteamina, tricarbóxiel fosfina (TCEP) o sales de ácido sulfuroso. Una vez obtenido, el ácido nucleico se desnatura por cualquier método conocido en la técnica para producir plantillas de ácido nucleico monocatenario y un par de primer y segundo oligonucleótidos se hibridan con la plantilla de ácido nucleico monocatenario de manera que el primer y segundo oligonucleótidos flanquean una región diana en la plantilla.

Los ácidos nucleicos pueden fragmentarse o romperse en fragmentos de ácido nucleico más pequeños. Los ácidos nucleicos, que incluyen ácidos nucleicos genómicos, pueden fragmentarse utilizando cualquiera de una variedad de métodos, tales como fragmentación mecánica, fragmentación química y fragmentación enzimática. Los métodos de fragmentación de ácido nucleico son conocidos en la técnica e incluyen, pero no se limitan a, digestión con DNasa, sonicación, cizallamiento mecánico y similares (J. Sambrook et al., "Molecular Cloning: A Laboratory Manual", 1989, 2.ª ed., Cold Spring Harbour Laboratory Press: New York, N.Y.; P. Tijssen, "Hybridization with Nucleic Acid Probes--Laboratory Techniques in Biochemistry and Molecular Biology (Parts I and II)", 1993, Elsevier; C. P. Ordahl et al., *Nucleic Acids Res.*, 1976, 3: 2985-2999; P. J. Oefner et al., *Nucleic Acids Res.*, 1996, 24: 3879-3889; Y. R. Thorstenson et al., *Genome Res.*, 1998, 8: 848-855). La publicación de patente US-2005/0112590 proporciona una visión general de diversos métodos de fragmentación conocidos en la técnica.

Los ácidos nucleicos genómicos pueden fragmentarse en fragmentos uniformes o fragmentarse aleatoriamente. Los ácidos nucleicos pueden fragmentarse para formar fragmentos que tienen una longitud de fragmento de aproximadamente 5 kilobases o 100 kilobases. Los fragmentos de ácido nucleico genómico pueden variar de 1 kilobases a 20 kilobases. Los fragmentos preferidos pueden variar en tamaño y tienen una longitud de fragmento promedio de aproximadamente 10 kilobases. Sin embargo, la longitud de fragmento deseada y los intervalos de longitudes de fragmentos pueden ajustarse dependiendo del tipo de dianas de ácido nucleico que se busca capturar. El método particular de fragmentación se selecciona para lograr la longitud de fragmento deseada. A continuación se proporcionan algunos ejemplos no limitantes.

La fragmentación química de los ácidos nucleicos genómicos puede lograrse utilizando varios métodos diferentes. Por ejemplo, las reacciones de hidrólisis que incluyen la base y la hidrólisis ácida son técnicas comunes utilizadas para fragmentar el ácido nucleico. La hidrólisis se facilita por aumentos de la temperatura, dependiendo del grado deseado de hidrólisis. La fragmentación puede lograrse alterando la temperatura y el pH como se describe a continuación. El beneficio de la hidrólisis basada en el pH para el cizallamiento es que puede dar como resultado productos monocatenarios. Además, la temperatura puede utilizarse con ciertos sistemas de amortiguador (p. ej.,

Tris) para desplazar temporalmente el pH hacia arriba o hacia abajo desde neutro para lograr la hidrólisis, a continuación de nuevo a neutro para almacenamiento a largo plazo, etc. Tanto el pH como la temperatura pueden modularse para afectar diferentes cantidades de cizallamiento (y, por lo tanto, distribuciones de longitud variable).

5 Otros métodos de fragmentación hidrolítica de ácidos nucleicos incluyen hidrólisis alcalina, fijación con formalina, hidrólisis por complejos metálicos (p. ej., porfirinas) y/o hidrólisis por radicales hidroxilo. Las láminas de ARN en condiciones alcalinas, véase, p. ej., Nordhoff et al., *Nucl. Acid. Res.*, 21 (15):3347-57 (2003), mientras que el ADN puede cortarse en presencia de ácidos fuertes.

10 Un protocolo de hidrólisis de ácido/base ilustrativo para producir fragmentos de ácido nucleico genómico se describe en Cargent et al. (1988) *Methods Enzymol.*, 152:432. Brevemente, se disuelven 1 g de ADN purificado en 50 mL de NaOH 0,1 N. Se añade 1,5 mL de HCl concentrado y la solución se mezcla rápidamente. El ADN precipitará inmediatamente y no debe agitarse durante más de unos pocos segundos para evitar la formación de un gran agregado. La muestra se incuba a temperatura ambiente durante 20 minutos para depurar parcialmente el ADN.
15 Posteriormente, se añaden 2 mL de NaOH 10 N (OH-concentración a 0,1 N) y la muestra se agita hasta que el ADN se vuelva a disolver completamente. A continuación, la muestra se incuba a 65 grados centígrados durante 30 minutos para hidrolizar el ADN. Los fragmentos resultantes típicamente varían desde aproximadamente 250-1000 nucleótidos pero pueden variar hacia abajo o hacia arriba dependiendo de las condiciones de la hidrólisis.

20 Después de que se haya purificado el ácido nucleico genómico, se vuelve a suspender en un amortiguador basado en Tris a un pH entre 7,5 y 8,0, tal como la solución hidratante de ADN de Qiagen. El ácido nucleico genómico que se ha vuelto a suspender se calienta a continuación a 65 °C y se incuba durante la noche. El calentamiento cambia el pH del amortiguador en el intervalo de bajo a medio 6, lo que conduce a la hidrólisis ácida. Con el tiempo, la hidrólisis ácida provoca que el ácido nucleico genómico se fragmente en productos monocatenarios y/o bicatenarios.

25 La escisión química también puede ser específica. Por ejemplo, las moléculas de ácido nucleico seleccionadas pueden escindirse mediante alquilación, particularmente moléculas de ácido nucleico modificadas con fosforotioato (véase, p. ej., K. A. Browne, "Metal ion-catalyzed nucleic Acid alkylation and fragmentation," *J. Am. Chem. Soc.* 124(27):7950-7962 (2002)). La alquilación en la modificación de fosforotioato hace que la molécula de ácido nucleico sea susceptible a la escisión en el sitio de modificación. Véase I. G. Gut and S. Beck, "A procedure for selective DNA alkylation and detection by mass spectrometry," *Nucl. Acids Res.* 23(8):1367-1373 (1995).

30 También se contempla el corte químico de ácidos nucleicos utilizando la técnica descrita en el método de secuenciación de Maxam-Gilbert (Chemical or Cleavage Method), *Proc. Natl. Acad. Sci. USA.* 74:560-564. En ese protocolo, el ácido nucleico genómico puede escindirse químicamente por exposición a productos químicos diseñados para fragmentar el ácido nucleico en bases específicas, tal como escisión preferencial a guanina, a adenina, a citosina y timina, y a citosina sola.

35 El cizallamiento mecánico de ácidos nucleicos en fragmentos puede ocurrir mediante la utilización de cualquier método conocido en la técnica. Por ejemplo, fragmentar los ácidos nucleicos puede lograrse mediante hidrocizallamiento, trituración a través de una aguja y sonicación. Véase, por ejemplo, Quail, et al. (Nov 2010) *DNA: Mechanical Breakage*. In: eLS. John Wiley & Sons, Chichester. doi:10.1002/9780470015902.a0005 333.pub2.

40 El ácido nucleico también puede cortarse mediante nebulización, véase (Roe, BA, Crabtree. JS and Khan, AS 1996); Sambrook & Russell, *Cold Spring Harb Protoc* 2006. La nebulización implica recoger ADN fragmentado de una neblina creada al forzar una solución de ácido nucleico a través de un pequeño orificio en un nebulizador. El tamaño de los fragmentos obtenidos por nebulización se determina principalmente por la velocidad a la que la solución de ADN pasa a través del orificio, alterando la presión del gas soplado a través del nebulizador, la viscosidad de la solución y la temperatura. Los fragmentos de ADN resultantes se distribuyen en un intervalo estrecho de tamaños (700-1330 pb). El deterioro de los ácidos nucleicos puede lograrse pasando ácidos nucleicos obtenidos a través de una capilaridad estrecha u orificio (Oefner et al., *Nucleic Acids Res.* 1996; Thorstenson et al., *Genome Res.* 1995). Esta técnica se basa en la hidrodinámica de puntos de punto que resulta cuando una muestra de ácido nucleico es forzada a través de un pequeño orificio mediante una bomba de jeringa.

45 En HydroShearing (Genomic Solutions, Ann Arbor, Mich., USA), el ADN en solución se hace pasar a través de un tubo con una contracción abrupta. A medida que se acerca a la contracción, el fluido acelera para mantener el caudal volumétrico a través del área más pequeña de la contracción. Durante esta aceleración, las fuerzas de arrastre estiran el ADN hasta que se brotan. Los fragmentos de ADN hasta que las piezas son demasiado cortas para las fuerzas de cizallamiento para romper los enlaces químicos. El caudal del fluido y el tamaño de la contracción determinan los tamaños finales del fragmento de ADN.

50 La sonicación también se utiliza para fragmentar ácidos nucleicos sometiendo el ácido nucleico a cortos períodos de sonicación, es decir, energía de ultrasonido. Un método de cizallamiento de ácidos nucleicos en fragmentos mediante sonicación se describe en la publicación de patente US-2009/0233814. En el método, se obtiene un ácido nucleico purificado colocado en una suspensión que tiene partículas dispuestas dentro. A continuación, la suspensión de la muestra y las partículas son sonicadas en fragmentos de ácido nucleico.

55

Un sistema basado en acústica que puede utilizarse para fragmentar el ADN está descrito en la patente US - 6.719.449, y la patente US-6.948.843 fabricado por Covares Inc. La patente US-6.235.501 describe un método de sonicación acústica mecánico para producir fragmentos de ADN de alto peso molecular mediante la aplicación de energía mecánica recíproca oscilante rápida en presencia de un medio líquido en un recipiente cerrado, que puede utilizarse para fragmentar mecánicamente el ADN.

Otro método de cizallamiento de ácidos nucleicos en fragmentos utiliza energía de ultrasonido para producir cavitación gaseosa en líquidos, tal como cizallamiento con BioRuptor de Diagoner (dispositivo de cizallamiento eléctrico, disponible comercialmente por Dienode, Inc.). La cavitación es la formación de pequeñas burbujas de gases o vapores disueltos debido a la alteración de la presión en los líquidos. Estas burbujas son capaces de vibración por resonancia y producen una turbulencia vigorosa o microcontinua. La tensión mecánica resultante puede conducir a cortar el ácido nucleico en fragmentos.

La fragmentación enzimática, también conocida como escisión enzimática, corta ácidos nucleicos en fragmentos utilizando enzimas, tales como endonucleasas, exonucleasas, ribozimas y ADNzimas. Dichas enzimas son ampliamente conocidas y están disponibles comercialmente, véase Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 3rd (2001) and Roberts RJ (January 1980). "Restriction and modification enzymes and their recognition sequences", *Nucleic Acids Res.* 8 (1): r63-r80. Las técnicas de fragmentación enzimática variables son bien conocidas en la técnica, y dichas técnicas se utilizan con frecuencia para fragmentar un ácido nucleico para secuenciación, por ejemplo, Aazard et al. 2002; Bentzley et al, 1998; Bentzley et al, 1996; Faulstich et al, 1997; Glover et al, 1995; Kirpekar et al, 1994; Owens et al, 1998; Pieles et al, 1993; Schuette et al, 1995; Smirnov et al, 1996; Wu & Aboleneen, 2001; Wu et al, 1998a.

Las enzimas más comunes utilizadas para fragmentar los ácidos nucleicos son endonucleasas. Las endonucleasas pueden ser específicas para una molécula de ácido nucleico bicatenario o monocatenaria. La escisión de la molécula de ácido nucleico puede producirse aleatoriamente dentro de la molécula de ácido nucleico o puede escindir en secuencias específicas de la molécula de ácido nucleico. La fragmentación específica de la molécula de ácido nucleico puede lograrse mediante la utilización de una o más enzimas en reacciones secuenciales o simultáneamente.

Las endonucleasas de restricción reconocen secuencias específicas dentro de ácidos nucleicos bicatenarios y generalmente escinden ambas cadenas dentro o cerca del sitio de reconocimiento para fragmentar el ácido nucleico. Las endonucleasas de restricción de origen natural se clasifican en cuatro grupos (tipos I, II III y IV) basados en sus requisitos de composición y cofactor enzimático, la naturaleza de su secuencia diana y la posición de su sitio de escisión de ADN en relación con la secuencia diana. Bickle TA, Krüger DH (June 1993), "Biology of DNA restriction," *Microbiol. Rev.* 57 (2): 434-50; Boyer HW (1971). "DNA restriction and modification mechanisms in bacteria". *Annu. Rev. Microbiol.* 25: 153-76; Yuan R (1981). "Structure and mechanism of multifunctional restriction endonucleases". *Annu. Rev. Biochem.* 50: 285-319. Todos los tipos de enzimas reconocen secuencias de ADN cortas específicas y llevan a cabo la escisión endonucleolítica de ADN para dar fragmentos específicos con fosfatos terminados en 5'. Las enzimas difieren en su secuencia de reconocimiento, composición de subunidades, posición de escisión y requisitos de cofactores. Williams RJ (2003). "Restriction endonucleases: classification, properties, and applications". *Mol. Biotechnol.* 23 (3): 225-43.

Cuando las endonucleasas de restricción reconocen secuencias específicas en ácidos nucleicos bicatenarios y generalmente escinden ambas cadenas, las endonucleasas de corte son capaces de escindir solo una de las cadenas del ácido nucleico en un fragmento. Las enzimas cortadas utilizadas para fragmentar ácidos nucleicos pueden ser naturales o modificadas genéticamente a partir de enzimas de restricción. Véase Chan et al., *Nucl. Acids Res.* (2011) 39 (1): 1-18.

El ADN puede romperse en procesos biológicos dentro de un organismo, o un medio biológico. Dicho ADN, o ADN libre de células, circula libremente en el torrente sanguíneo. Por ejemplo, el ADN fetal libre de células (ADNc) es ADN fetal que circula libremente en el torrente sanguíneo materno. El ADN tumoral libre de células (ADNc) es ADN tumoral que circula libremente en el torrente sanguíneo. Puede utilizarse ADN fragmentado o cortado, sin embargo, el ADN se obtiene de forma fragmentada.

En realizaciones preferidas de la presente invención, los fragmentos de ácido nucleico se unen entre sí en un complejo, por ejemplo, véase la Figura 3 para fragmentos idénticos y la Figura 16 para dos cadenas de ácido nucleico dúplex. Puede utilizarse cualquier molécula de enlace para unir las moléculas. El enlazador utilizado en la presente invención puede sintetizarse u obtenerse comercialmente de varias compañías, por ejemplo, Integrated DNA Technologies, Inc., Gene Link, Inc., y TriLink Biotechnologies, Inc. El enlazador puede ser cualquier molécula para unir dos cebadores o dos fragmentos de ácido nucleico. La molécula de enlace también puede unirse a múltiples fragmentos juntos. Puede incorporarse cualquier número de fragmentos al complejo.

En ciertas realizaciones, la molécula de enlace puede ser una molécula de estreptavidina y los fragmentos a unir pueden comprender ácido nucleico biotinilado. En realizaciones en las que se utilizan cebadores enlazados para crear los fragmentos de ácido nucleico enlazados mediante amplificación, los cebadores pueden biotinilarse y unirse entre sí en una molécula de estreptavidina. Por ejemplo, 4 fragmentos pueden unirse entre sí en una estreptavidina de tetrámero. Se podrían unir más de cuatro moléculas a través de la formación de concatámeros, por ejemplo. En ciertos métodos de la invención, dos o más fragmentos de ácido nucleico pueden unirse a través de reacciones de química clic. Véase Kolb, et al., *Click Chemistry: Diverse Chemical Function from a Few Good Reactions*, *Angew Chem Int Ed Engl.* 2001 Jun 1;40(11):2004-2021.

Las moléculas de enlace, por ejemplo y de varias nanopartículas conocidas, pueden enlazar un gran número de fragmentos que incluyen cientos o miles de fragmentos en una sola molécula enlazada. Un ejemplo de una nanopartícula de unión puede ser nanopartículas de oro de ADN polivalente que comprenden oro coloidal modificado con secuencias de ADN sintéticas con caperuza de tiol en su superficie. Véase Mirkin, et al., 1996, A DNA-based method for rationally assembling nanoparticles into macroscopic materials, *Nature*, 382:607-609. Las secuencias de ADN de superficie pueden ser complementarias a las secuencias de moléculas plantilla deseadas o pueden comprender cebadores universales.

La molécula de enlace también puede servir para separar los fragmentos de ácido nucleico. En realizaciones preferidas, los fragmentos están orientados para evitar la unión entre ellos. Con el enlazador que crea separación espacial y orientación de los fragmentos controlados, puede impedirse el colapso o la unión entre los fragmentos.

Los enlazadores pueden ser polietilenglicol (PEG) o un PEG modificado. Un PEG modificado, tal como DBCO-PEG₄ o PEG-11 puede utilizarse para unir los dos adaptadores o ácidos nucleicos. En otro ejemplo, se utiliza PEG modificado con N-hidroxisuccinimida (NHS) para unir los dos adaptadores. Véase Schlingman, et al., *Colloids and Surfaces B: Biointerfaces* 83 (2011) 91-95. Puede utilizarse cualquier oligonucleótido u otra molécula para unir adaptadores o ácidos nucleicos.

Los aptámeros pueden utilizarse para unir dos adaptadores o ácidos nucleicos. Los aptámeros pueden diseñarse para unirse a diversas dianas moleculares, tales como cebadores o ácidos nucleicos. Los aptámeros pueden diseñarse o seleccionarse por el SELEX (evolución sistemática de ligandos mediante enriquecimiento exponencial). Los aptámeros son macromoléculas de ácido nucleico que se unen específicamente a moléculas diana. Al igual que todos los ácidos nucleicos, un ligando de ácido nucleico particular, es decir, un aptámero, puede describirse mediante una secuencia lineal de nucleótidos (A, U, T, C y G), típicamente 15-40 nucleótidos de longitud. Los aptámeros pueden incluir bases invertidas o bases modificadas. Los aptámeros o aptámeros modificados pueden incluir al menos una base invertida o base modificada.

Debe apreciarse que el enlazador puede estar compuesto de bases invertidas, o comprender al menos una base invertida. Las bases invertidas o bases modificadas pueden adquirirse a través de cualquier entidad comercial. Las bases invertidas o bases modificadas están desarrolladas y disponibles comercialmente. Las bases invertidas o bases modificadas pueden incorporarse en otras moléculas. Por ejemplo, la 2-aminopurina puede sustituirse en un oligonucleótido. La 2-aminopurina es una base fluorescente que es útil como sonda para controlar la estructura y la dinámica del ADN. La 2,6-diaminopurina (2-Amino-dA) es una base modificada que puede formar tres enlaces de hidrógeno cuando se empareja con base con dT y puede aumentar la T_m de los oligos cortos. 5-bromo-desoxiuridina es una base halogenada fotorreactiva que puede incorporarse en oligonucleótidos para reticularlas a ADN, ARN o proteínas con exposición a luz UV. Otros ejemplos de bases invertidas o bases modificadas incluyen desoxiuridina (dU), dT invertida, didesoxicitidina (ddC), 5-metil desoxicitidina o 2'-desoxiinosina (dI). Debe apreciarse que cualquier base invertida o modificada puede utilizarse para unir ácidos nucleicos plantilla.

En realizaciones preferidas, el enlazador comprende una molécula para unir dos cebadores o dos fragmentos de ácido nucleico. El enlazador puede ser una sola molécula o una pluralidad de moléculas. El enlazador puede comprender unas pocas bases invertidas o bases modificadas, o bases completamente invertidas o bases modificadas. El enlazador puede comprender tanto bases de Watson-Crick como bases invertidas o modificadas.

Debe apreciarse que puede utilizarse cualquier molécula espaciadora o molécula de unión en la presente invención. El enlazador o molécula espaciadora puede ser un lípido o un oligosacárido, o un oligosacárido y un lípido. Véase la patente US-5.122.450. En este ejemplo, la molécula es preferiblemente una molécula lipídica y, más preferiblemente, un glicérido o fosfano que posee al menos dos cadenas de polialquileno hidrófobas.

El enlazador puede estar compuesto por cualquier número de adaptadores, cebadores y copias de fragmentos. Un enlazador puede incluir dos brazos idénticos, donde cada brazo está compuesto por moléculas de unión, cebadores de amplificación, cebadores de secuenciación, adaptadores y fragmentos. Un enlazador puede unir cualquier número de brazos, como tres o cuatro brazos. Las plantillas de ácido nucleico pueden estar enlazadas por una molécula espaciadora. El enlazador en la presente invención puede ser cualquier molécula o método para unir dos fragmentos o cebadores. Polietilenglicol o un PEG modificado tal como DBCO-PEG₄ o PEG-11. El enlazador puede ser un lípido o un hidrocarburo. Alternativamente, una proteína puede unirse a los adaptadores o los ácidos nucleicos. Un oligosacárido puede enlazar los cebadores o ácidos nucleicos o los aptámeros pueden enlazar los cebadores o ácidos nucleicos. Cuando los fragmentos están enlazados, las copias están orientadas para estar en fase para evitar la unión entre ellas.

Un enlazador puede ser un anticuerpo. El anticuerpo puede ser un monómero, un dímero o un pentámero. Debe apreciarse que puede utilizarse cualquier anticuerpo para unir dos cebadores o ácidos nucleicos. Por ejemplo, se conoce en la técnica que el nucleósido puede hacerse inmunogénico mediante el acoplamiento a proteínas. Véase Void, BS (1979), *Nucl Acids Res* 7, 193-204. Además, los anticuerpos pueden prepararse para unirse a ácidos nucleicos modificados. Véase *Biochemical Education*, vol. 12, edición 3.

El enlazador puede permanecer unido al complejo durante la amplificación. Alternativamente, el enlazador puede eliminarse antes de la amplificación. Un enlazador puede unirse a una molécula de unión, y la molécula de unión se une a un cebador de amplificación. Cuando se elimina el enlazador, la molécula de unión o el cebador de unión

se expone. La molécula de unión expuesta también se une a un soporte sólido y se forma un arco. El enlazador puede eliminarse mediante cualquier método conocido en la técnica, incluyendo lavar con un disolvente, aplicar calor, alterar el pH, lavar con un detergente o tensioactivo, etc.

5 Los métodos de la invención proporcionan ácidos nucleicos para enlazarse junto con una molécula enlazadora. En muestras con material genético bajo, los ácidos nucleicos pueden unirse entre sí para asegurar que fragmentos idénticos o fragmentos que comprenden ambas cadenas de un ácido nucleico dúplex se amplifican simultáneamente o secuencialmente. Las muestras tales como las muestras prenatales tienen un bajo contenido genético y la amplificación de fragmentos enlazados de acuerdo con la invención aumenta el contenido detectable.
10 Este método reduce la relación señal/ruido, mejorando la detección de la secuencia diana.

Los métodos de la invención utilizan la amplificación para amplificar un ácido nucleico diana, tal como un fragmento, a un nivel detectable. En la presente invención se utiliza amplificación por PCR. Además, los segmentos amplificados creados por un proceso de amplificación pueden ser mismos, plantillas eficientes para amplificaciones posteriores.

15 La amplificación se refiere a la producción de copias adicionales de una secuencia de ácido nucleico Y generalmente se lleva a cabo utilizando reacción en cadena de la polimerasa u otras tecnologías bien conocidas en la técnica (p. ej., Dieffenbach and Dveksler, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. [1995]). La reacción de amplificación puede ser cualquier reacción de amplificación conocida en la técnica que amplifica las moléculas de ácido nucleico, tales como la reacción en cadena de la polimerasa, la reacción en cadena de la polimerasa anidada, la reacción en cadena de la ligasa (Barany F. (1991) PNAS 88:189-193; Barany F. (1991) PCR Methods and Applications 1:5-16), ligase detection reaction (Barany F. (1991) PNAS 88:189-193), sistema de amplificación basado en transcripción, amplificación basada en secuencia de ácido nucleico, amplificación por círculo rodante y amplificación por círculo rodante hiperramificado.

25 Amplificación por desplazamiento múltiple (MDA), una técnica de amplificación de ADN que no está basada en PCR, amplifica rápidamente cantidades mínimas de muestras de ADN para el análisis genómico. La reacción comienza por hibridación de cebadores hexámeros aleatorios a la plantilla: La síntesis de ADN se lleva a cabo mediante una enzima de alta fidelidad a una temperatura constante.

30 En la invención, la reacción de amplificación es la reacción en cadena de la polimerasa. La reacción en cadena de la polimerasa (PCR) se refiere a los métodos de K. B. Mullis (patentes US-4.683.195 y 4.683.202) para aumentar la concentración de un segmento de una secuencia diana en una mezcla de ADN genómico sin clonación o purificación. El proceso para amplificar la secuencia diana incluye introducir un exceso de cebadores oligonucleotídicos a una mezcla de ADN que contiene una secuencia diana deseada, seguido de una secuencia precisa de ciclado térmico en presencia de una ADN polimerasa. Los cebadores son complementarios a sus respectivas cadenas de la secuencia diana bicatenaria.

35 En algunos aspectos de la invención, los cebadores de PCR están unidos por una molécula de enlace y a través del proceso de PCR, las copias idénticas de un fragmento o ambas cadenas de un fragmento dúplex están enlazadas a los cebadores. En otras realizaciones, se añaden adaptadores a los cebadores o copias de los fragmentos. El complejo resultante incluye, generalmente, dos copias idénticas de un fragmento o las cadenas sentido y antisentido de un ácido nucleico dúplex unido directa o indirectamente por una molécula de enlace. Debe apreciarse que uno o ambos fragmentos o cadenas enlazados pueden incluir un error. Sin embargo, existe una baja probabilidad de que cada uno tenga un error de coincidencia en la misma base exacta. El desacuerdo entre los dos fragmentos en una base indicaría un error en lugar de una variante verdadera. La base podría identificarse como un desconocido, solo a partir de los datos de secuenciación sin procesar.

45 Los cebadores pueden prepararse mediante una variedad de métodos que incluyen, pero no se limitan a, clonación de secuencias apropiadas y síntesis química directa mediante la utilización de métodos bien conocidos en la técnica (Narang et al., Methods Enzymol., 68:90 (1979); Brown et al., Methods Enzymol., 68:109 (1979)). Los cebadores también pueden obtenerse de fuentes comerciales tales como Operon Technologies, Amersham Pharmacia Biotech, Sigma, y Life Technologies. Los cebadores pueden tener una temperatura de fusión idéntica. Las longitudes de los cebadores pueden extenderse o acortarse en el extremo 5' o en el extremo 3' para producir cebadores a las temperaturas de fusión deseadas. Además, la posición de hibridación de cada par de cebadores puede diseñarse de manera que la secuencia y la longitud de los pares de cebadores produzcan la temperatura de fusión deseada. La ecuación más simple para determinar la temperatura de fusión de los cebadores más pequeños que 25 pares de bases es la Regla de Wallace ($T_d = 2(A + T) + 4(G + C)$). Los programas informáticos también pueden utilizarse para diseñar cebadores, que incluyen, pero no se limitan a, Array Designer Software (Arrayit Inc.), Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), NetPrimer, y DNAsis de Hitachi Software Engineering. La T_M (temperatura de fusión o de hibridación) de cada cebador se calcula utilizando programas de software tales como Oligo Design, disponible en Invitrogen Corp.

60 En algunas realizaciones, para efectuar la amplificación, se desnaturaliza una mezcla y a continuación los cebadores se hibridan con sus secuencias complementarias dentro de la molécula diana. Después de la hibridación, los cebadores se extienden con una polimerasa para formar un nuevo par de cadenas complementarias. Las etapas de desnaturalización, hibridación del cebador y extensión de la polimerasa pueden repetirse muchas veces (es decir, desnaturalización, hibridación y extensión constituyen un ciclo; puede haber numerosos ciclos) para obtener una alta concentración de un segmento amplificado de una secuencia diana deseada. La longitud del segmento amplificado de la secuencia diana deseada se determina por posiciones relativas de los cebadores entre sí y, por lo tanto, esta longitud es un parámetro controlable.

65

5 En algunas realizaciones, para crear complejos de la invención, los cebadores están enlazados por una molécula de unión o una molécula espaciadora para crear dos copias enlazadas del fragmento. En otras realizaciones, dos fragmentos se enlazan entre sí después de al menos una etapa de PCR. Debe apreciarse que la PCR puede aplicarse a fragmentos antes o después de que los fragmentos se unan mediante una molécula de enlace. En algunas realizaciones, cuando los fragmentos se unen, la PCR puede implementarse en los fragmentos enlazados. En algunas realizaciones, las copias enlazadas experimentan amplificación. La etapa de amplificación incluye cebadores enlazados. El resultado es que después de un ciclo de PCR, se producen complejos enlazados que comprenden copias de los fragmentos.

10 Con la PCR, es posible amplificar una sola copia de una secuencia diana específica en el ADN genómico a un nivel que puede detectarse mediante varias metodologías diferentes (p. ej., tinción, hibridación con una sonda marcada; incorporación de cebadores biotinilados seguido de detección del conjugado de la enzima avidina-enzima; incorporación de trifosfatos de desoxinucleótidos marcados con ³²P, tales como dCTP o dATP, en el segmento amplificado). Además del ADN genómico, cualquier secuencia de oligonucleótidos puede amplificarse con el conjunto apropiado de moléculas de cebador. En particular, los segmentos amplificados creados por el propio proceso de PCR son, por sí mismos, plantillas eficientes para las posteriores amplificaciones por PCR. Las secuencias diana amplificadas pueden utilizarse para obtener segmentos de ADN (p. ej., genes) para su inserción en vectores recombinantes.

20 Otros métodos de amplificación y estrategias se describen en la presente memoria. Por ejemplo, otro enfoque sería combinar la PCR y la reacción en cadena de la ligasa (LCR). Dado que la PCR amplifica más rápido que la LCR y requiere menos copias de ADN diana para iniciar, la PCR puede utilizarse como primera etapa seguida de LCR. El producto amplificado se podría utilizar en una reacción de detección de LCR o ligasa (LDR) de una manera específica de alelo que indicaría si estaba presente una mutación. Otro enfoque es utilizar LCR o LDR tanto para la amplificación como para la discriminación específica de alelo. La reacción posterior es ventajosa porque da como resultado una amplificación lineal. Por lo tanto, la cantidad de producto amplificado es un reflejo de la cantidad de ADN diana en la muestra original y, por lo tanto, permite la cuantificación.

30 La LCR utiliza pares de oligonucleótidos adyacentes que son complementarios a la longitud completa de la secuencia diana (Barany F. (1991) PNAS 88:189-193; Barany F. (1991) PCR Methods and Applications 1:5-16). Si la secuencia diana es perfectamente complementaria a los cebadores en la unión de estas secuencias, una ADN ligasa enlazará los nucleótidos terminales adyacentes 3' y 5' que forman una secuencia combinada. Si se utiliza una ligasa de ADN termoestable con ciclos térmicos, la secuencia combinada se amplificará secuencialmente. Una única falta de coincidencia de base en la unión de los oligonucleótidos evitará la ligación y la amplificación. Por lo tanto, el proceso es alelo específico. Otro conjunto de oligonucleótidos con nucleótidos 3'específicos para el mutante se utilizaría en otra reacción para identificar el alelo mutante. Se podría utilizar una serie de condiciones estándar para detectar todas las posibles mutaciones en cualquier sitio conocido. La LCR utiliza típicamente tanto cadenas de ADN genómico como dianas para la hibridación de oligonucleótidos con cuatro cebadores, y el producto se aumenta exponencialmente mediante ciclos térmicos repetidos.

40 Los adaptadores o códigos de barras de amplificación o secuenciación, o una combinación de los mismos, pueden unirse al ácido nucleico fragmentado. Dichas moléculas pueden obtenerse comercialmente, tal como de Integrated DNA Technologies (Coralville, IA). En ciertas realizaciones, dichas secuencias se unen a la molécula de ácido nucleico plantilla con una enzima tal como una ligasa. Las ligasas adecuadas incluyen ADN ligasa de T4 y ligasa de ARN T4, disponibles comercialmente de New England Biolabs (Ipswich, MA). La ligación puede tener extremos romos o llevarse a cabo mediante la utilización de extremos sobresalientes complementarios.

45 Después de la fragmentación, los extremos de los fragmentos pueden repararse, recortarse (p. ej., utilizando una exonucleasa), o rellenarse (p. ej., utilizando una polimerasa y dNTP) para formar extremos romos. La reparación final puede realizarse para generar extremos romos de ácido nucleico fosforilados de 5' utilizando kits comerciales, tales como los disponibles en Epicentre Biotechnologies (Madison, WI). Tras generar extremos romos, los extremos pueden tratarse con una polimerasa y dATP para formar una adición independiente de plantilla hasta el extremo 3'y el extremo 5' de los fragmentos, produciendo así un único saliente A. Este único A puede guiar el ligamiento de fragmentos con una sola T que sobresale desde el extremo 5' en un método denominado clonación T-A. Alternativamente, debido a que la posible combinación de sobresalientes que quedan en las enzimas de restricción se conoce después de una digestión de restricción, los extremos pueden dejarse como, es decir, extremos irregulares. Pueden utilizarse oligonucleótidos bicatenarios con extremos sobresalientes complementarios.

55 En ciertas realizaciones, uno o más códigos de barras están unidos a cada uno, cualquiera o todos los fragmentos. Una secuencia de código de barras generalmente incluye ciertas características que hacen que la secuencia sea útil en las reacciones de secuenciación. Las secuencias de código de barras están diseñadas de tal forma que cada secuencia esté correlacionada con una porción particular de ácido nucleico, lo que permite correlacionar las lecturas de secuencias con la porción de la que proceden. Los métodos de diseño de conjuntos de secuencias de código de barras se muestran, por ejemplo, en la patente US- 6.235.475. En ciertas realizaciones, las secuencias de código de barras están unidas a la molécula de ácido nucleico plantilla, p. ej., con una enzima. La enzima puede ser una ligasa o una polimerasa, como se discutió anteriormente. Unir secuencias de código de barras a plantillas de ácido nucleico se muestra en la publicación US- 2008/0081330 y publicación US- 2011/0301042. Los métodos para diseñar conjuntos de secuencias de código de barras y otros métodos para unir secuencias de código de barras se muestran en las páginas de las patentes US-

7.537.897; 6.138.077; 6.352.828; 5.636.400; 6.172.214; y 5.863.722. Después de cualquier etapa de procesamiento (p. ej., obtención, aislamiento, fragmentación, amplificación o código de barras), puede secuenciarse ácido nucleico.

5 Los métodos ilustrativos para diseñar conjuntos de secuencias de código de barras y otros métodos para unir secuencias de código de barras se muestran en los números de patentes US-6.138.077; 6.352.828; 5.636.400; 6.172.214; 6.235.475; 7.393.665; 7.544.473; 5.846.719; 5.695.934; 5.604.097; 6.150.516; RE39.793; 7.537.897; 6.172,218; y 5.863.722.

10 La secuencia del código de barras incluye generalmente ciertas características que hacen que la secuencia sea útil en las reacciones de secuenciación. Por ejemplo, las secuencias de código de barras pueden diseñarse para tener regiones de homopolímero mínima o nula, es decir, 2 o más de la misma base en una fila tal como AA o CCC, dentro de la secuencia de código de barras. Las secuencias de código de barras también pueden diseñarse de modo que no solapen la región diana para ser una secuencia o contengan una secuencia que sea idéntica a la diana.

15 Las secuencias de código de barras primera y segunda se diseñan de tal forma que cada par de secuencias se correlacione con una muestra particular, lo que permite distinguir y validar las muestras. Los métodos para diseñar conjuntos de secuencias de código de barras se muestran, por ejemplo, en Brenner et al. (patente US-6.235.475). Las secuencias de código de barras pueden variar de aproximadamente 2 nucleótidos a aproximadamente 50; y preferiblemente de aproximadamente 4 a aproximadamente 20 nucleótidos. Dado que la secuencia del código de barras se secuencia junto con el ácido nucleico plantilla o puede secuenciarse en una lectura separada, la longitud del oligonucleótido debe tener una longitud mínima para permitir la lectura más larga del ácido nucleico plantilla unido. Generalmente, las secuencias de código de barras están separadas de la molécula de ácido nucleico plantilla por al menos una base.

25 Los métodos de la invención pueden implicar unir las secuencias de código de barras a los ácidos nucleicos plantilla. Los ácidos nucleicos plantilla pueden fragmentarse o romperse a la longitud deseada, p. ej., generalmente de 100 a 500 bases o más, mediante la utilización de una variedad de métodos mecánicos, químicos y/o enzimáticos. El ADN puede cortarse aleatoriamente mediante sonicación, expuesto a una ADNasa o una o más enzimas de restricción, una transposasa o enzima de corte. El ARN puede fragmentarse mediante una breve exposición a una ARNasa, calor más magnesio o mediante cizallamiento. El ARN puede convertirse en ADNc antes o después de la fragmentación.

30 La secuencia de código de barras se integra con la plantilla utilizando métodos conocidos en la técnica. La secuencia de código de barras se integra con la plantilla utilizando, por ejemplo, una ligasa, una polimerasa, clonación Topo (p. ej., sistema de clonación de vector topoisomerasa de Invitrogen utilizando una enzima topoisomerasa) o ligamiento o conjugación química. La ligasa puede ser cualquier enzima capaz de ligar un oligonucleótido (ARN o ADN) a la molécula de ácido nucleico plantilla. Las ligasas adecuadas incluyen ADN ligasa de T4 y ARN ligasa de T4 (tales ligasas están disponibles comercialmente, en New England Biolabs). Los métodos para utilizar ligasas son bien conocidos en la técnica. La polimerasa puede ser cualquier enzima capaz de añadir nucleótidos al extremo 3' y 5' de las moléculas de ácido nucleico de la plantilla. La secuencia de código de barras puede incorporarse a través de una reacción de PCR como parte del cebador de PCR. Independientemente de la incorporación de códigos de barras moleculares o la ubicación de los códigos de barras en el caso de que se incorporen, se pueden unir adaptadores de secuenciación al producto de ácido nucleico en forma bidireccional, de manera que en la misma ejecución de secuenciación habrá lecturas de secuenciación desde ambos extremos 5' y 3' de la secuencia diana. En algunos casos, es ventajoso utilizar la ubicación del código de barras en el extremo 5' o 3' de la secuencia diana para indicar la dirección de la lectura. Es bien conocido por un experto en la técnica cómo unir los adaptadores de secuenciación utilizando técnicas tales como PCR o ligamiento.

45 La Figura 6 muestra ejemplos de posibles configuraciones de adaptador y cebadores. Como se muestra en 602, un cebador P7 está unido a un sitio de cebador Read2, que está unido a una región complementaria. En 603, una región de cebado enlazada por PCR se une a un identificador molecular único. Como se muestra en 604, un cebador P5 se une a un sitio de cebador de lectura de índice y un sitio de control de siembra.

50 En algunas realizaciones, se unen múltiples copias de un fragmento entre sí. Debe apreciarse que cualquier número de fragmentos puede unirse entre sí, ya sea 2, 3, 4, etc. Las copias unidas pueden denominarse unidad. A continuación, varias unidades pueden unirse junto con una molécula de enlace. Debe apreciarse que cualquier número de unidades puede estar unida por una molécula de enlace. Esto aumenta la densidad de información dentro de un complejo. Cuando el complejo se une a un soporte sólido, el complejo se amplifica. Los productos de amplificación pueden unirse al soporte sólido. Al unir múltiples copias del fragmento al complejo y a continuación amplificar los complejos, aumenta la densidad de información sobre un soporte sólido.

60 Los ácidos nucleicos pueden amplificarse mediante dos o más cebadores unidos. Puede utilizarse cualquier método de amplificación conocido junto con los cebadores enlazados. La PCR Digital o la PCR en emulsión pueden utilizarse para crear dos o más fragmentos de ácido nucleico enlazados para sembrar los agrupamientos de secuenciación o para su utilización en otros métodos de secuenciación. Puede crearse un ácido nucleico plantilla ligando adaptadores a un fragmento de ácido nucleico de interés a secuenciar. Los adaptadores pueden incluir opcionalmente sitios de cebado universales, uno o más sitios de cebador de secuenciación e identificadores de agrupamiento únicos para garantizar que todas las lecturas de secuenciación en un agrupamiento dado se originaron a partir de la misma plantilla de inicio. La Figura 9 muestra adaptadores de ligamiento ilustrativos. Por ejemplo, los adaptadores pueden utilizarse con regiones madre variables tales

65

como y1: CCTTCTGCTAC (ID. DE SEC. N.º: 1), y2: ATGCGAGCCTCT (ID. DE SEC. N.º: 2), y3: GCACCTCATCCA (ID. DE SEC. N.º: 3), e y4: TGCAGGGGGTG (ID. DE SEC. N.º: 4). Las secuencias del adaptador pueden incluir un identificador de agrupamiento único (UCI) que puede comprender una serie de bases aleatorias (p. ej., 2, 3, 4, 5 o más) para distinguir entre agrupamientos vecinos en una celda de flujo de secuenciación. Las secuencias adaptadoras pueden incluir un T unido a fosforotioato para reducir la digestión con exonucleasa 3' que podría eliminar el saliente T y reducir la eficiencia de la ligación. Un bloqueador de fosfato 3' es opcional pero no esencial para los métodos de PCR digital de la invención.

Una vez que los adaptadores se han ligado al fragmento de ácido nucleico a secuenciar, se puede crear una emulsión o gota. Las gotas pueden ser gotas acuosas rodeadas por un fluido portador inmiscible. Los métodos para formar tales gotas y llevar a cabo la amplificación por PCR dentro de las gotas se muestran, por ejemplo, en Link et al. (solicitudes de patentes US-2008/0014589, 2008/0003142, and 2010/0137163), Stone et al. (patente US- 7.708.949 y solicitud de patente US-2010/0172803), and Anderson et al. (patente US- 7.041.481 y que fue reexpedida como RE41,780).

Puede añadirse una única molécula de ácido nucleico plantilla a una gota para garantizar que los agrupamientos eventuales se siembren con solo una molécula plantilla y para evitar la formación de complejos de ácido nucleico enlazados híbridos que contengan diferentes fragmentos de ácido nucleico. También se añaden varios cebadores múltiplex que pueden ser específicos del gen a la gota junto con cebadores unidos. Los cebadores enlazados pueden ser dos o más cebadores enlazados entre sí de acuerdo con cualquiera de los métodos descritos en la presente memoria. Los cebadores enlazados pueden incluir, por ejemplo, sitios de cebado universales correspondientes a los sitios de cebado universales en los adaptadores ligados, así como también sitios de cebadores de secuenciación (p. ej., sitios de cebado de índice diferentes para identificar cuando más de una molécula ha sembrado un agrupamiento). Los cebadores enlazados pueden incluir cebadores específicos de genes dirigidos a regiones específicas de interés a secuenciar de manera que pueda evitarse la etapa de ligamiento inicial y puede añadirse un fragmento de ácido nucleico no modificado directamente a la gota para la amplificación por PCR digital de cebador enlazado. La Figura 10 ilustra cebadores y fragmentos de ácido nucleico plantilla ligados donde la plantilla ligada comprende un sitio de cebado correspondiente al sitio de cebado de los cebadores universales enlazados y cebadores múltiplex específicos de genes se utilizan para crear copias enlazadas del ácido nucleico diana.

La Figura 7 ilustra métodos de ejemplo de PCR de emulsión de cebador enlazado. Un fragmento de ácido nucleico de muestra 701, tal como ADN libre de células que tiene una región de interés 724, se liga 703 con adaptadores 725 para formar una molécula plantilla 730. A continuación se crea una emulsión 705 que comprende una sola molécula plantilla 730, cebadores unidos 722 y cebadores específicos de genes múltiplex 723 en una gota 726. A continuación, la plantilla se amplifica en la gota utilizando métodos de PCR de emulsión conocidos para crear copias enlazadas de la molécula de plantilla 730 que comprenden sitios de cebadores de secuenciación, la región de ácido nucleico de interés 724 y cualquiera de las otras secuencias opcionales descritas en la presente memoria. A continuación, la emulsión se rompe en 707 de acuerdo con cualquier método conocido para liberar las copias enlazadas de la molécula de plantilla. Cabe señalar que la amplificación por PCR puede introducir errores de PCR 721 en las moléculas plantilla enlazadas, pero la naturaleza de los métodos descritos con siembra dual de agrupamiento es que dichos errores pueden identificarse y diferenciarse de las variantes verdaderas 720 que estarían presentes en ambas copias enlazadas que siembran un agrupamiento. Las moléculas plantilla 730 enlazadas pueden a continuación seleccionarse, purificarse o seleccionarse enzimáticamente en 709 para la siembra de agrupamientos de secuenciación. Las plantillas 730 enlazadas pueden a continuación hibridarse con una celda de flujo o secuenciarse de otra manera en 711 como se describe en otra parte.

Como se señaló anteriormente, las moléculas de plantilla enlazadas pueden comprender dos o más copias de un fragmento de ácido nucleico. La Figura 8 ilustra una gota 726 utilizada para la formación de moléculas plantilla enlazadas o complejas que comprenden cuatro copias de fragmentos. La gota 726 comprende complejos de cuatro cebadores 722 unidos con dos copias de dos cebadores alternos. Dicho complejo de cebadores unidos puede formarse utilizando, por ejemplo, tetrámeros de estreptavidina/biotina como se describió anteriormente. La gota 726 puede comprender además cebadores múltiplex específicos de genes 723 y, preferentemente, una sola copia de la molécula de plantilla de fragmento de ácido nucleico 730 a secuenciar. La gota puede someterse de otra manera a amplificación por PCR de la misma manera que se describe con respecto a la Figura 7 con el resultado alternativo de producir cuatro copias enlazadas de la molécula de plantilla con las que sembrar un agrupamiento para secuenciación y ayudar en la identificación y diferenciación de las variantes legítimas, si raras, 720 y errores de PCR 721 u otros falsos positivos.

Se describen métodos que se refieren a reducir la interferencia entre agrupamientos cercanos en corridas de secuenciación de alta densidad. La Figura 15 ilustra el problema presentado por interferencia en el que una base coincidente a partir de un agrupamiento vecino durante una lectura de secuenciación puede enmascarar una base mixta verdadera dentro de un solo agrupamiento. Para reducir esta interferencia, puede añadirse un segundo adaptador a un sitio de cebador de secuenciación diferente durante el ligamiento y utilizarse un cebador de secuenciación correspondiente para ayudar a diferenciar entre el acuerdo de bases dentro de un agrupamiento y el acuerdo de bases entre agrupamientos vecinos. Si bien estos métodos pueden utilizarse en cualquier técnica de secuenciación estándar, las Figuras 13 y 14 ilustran los métodos aplicados a PCR de emulsión de cebador enlazado y siembra de agrupamiento de fragmentos enlazados. La Figura 13 muestra múltiples fragmentos de ácido nucleico de muestra 701 y 735, tales como ADN libre de células que tiene regiones 724 y 736 de interés respectivamente, que se ligan 703 con dos adaptadores 725 y 734 diferentes para formar una primera molécula plantilla 730 y una segunda molécula plantilla 733, que comprende las diferentes regiones 724 y 736 de interés junto con las variantes 720 a caracterizar. A continuación, se crean dos emulsiones 705, comprendiendo cada una de las moléculas de plantilla 730

o 733, cebadores unidos 722, correspondientes a cada adaptador, y cebadores específicos de genes múltiple 723 en una gota 726. A continuación, las plantillas se amplifican en la gota utilizando métodos de PCR de emulsión conocidos para crear copias enlazadas de las moléculas de plantilla 730 y 733 que comprenden diferentes sitios de cebadores de secuenciación introducidos por los dos adaptadores diferentes 725 y 734. A continuación, la emulsión se rompe 707 de acuerdo con cualquier método conocido para liberar las copias enlazadas de las moléculas de plantilla 730 y 733. Cabe señalar que la amplificación por PCR puede introducir errores de PCR 721 en las moléculas de plantilla enlazadas, pero la naturaleza de los métodos descritos con siembra dual de agrupamiento es que dichos errores pueden identificarse y diferenciarse de las variantes verdaderas 720 que estarían presentes en ambas copias enlazadas que siembran un agrupamiento. Las moléculas de plantilla 730 y 733 enlazadas pueden a continuación seleccionarse, purificarse o seleccionarse enzimáticamente 709 para la siembra de agrupamientos de secuenciación. Las plantillas 730 enlazadas pueden a continuación hibridarse con una celda de flujo o secuenciarse de otra manera en 711 como se describe en otra parte.

Como se muestra en la Figura 14, los agrupamientos se siembran con uno de los dos adaptadores de plantillas 725 o 734. Las etapas ilustrativas del método de secuenciación incluyen a continuación realizar una lectura corta utilizando un sitio de cebado de secuenciación común en un sentido de todos los agrupamientos; realizar una lectura de índice común en el otro sentido de todos los agrupamientos para identificar la posible siembra del agrupamiento por plantillas múltiples; realizar una lectura en el agrupamiento diana con cebadores correspondientes al sitio de cebado de secuenciación del segundo adaptador 734; y a continuación realizar una lectura del índice largo en el agrupamiento vecino con cebadores correspondientes al sitio de cebado de secuenciación del primer adaptador 725. Hay una probabilidad del 50 % de que los agrupamientos vecinos contengan el sitio de cebado de secuenciación del primer adaptador 725 y, mediante la utilización de la técnica anterior, pueden identificarse las lecturas de agrupamientos vecinos y se reduce el riesgo de que las bases mixtas de enmascaramiento de interferencias dentro de un solo agrupamiento. La identificación de las variantes 720 y su diferenciación de los errores de PCR 721, puede lograrse con mayor fiabilidad gracias a la menor diafonía que ofrecen los métodos descritos.

Como se muestra en la Figura 1A, un enlazador 103 comprende dos cebadores cortos 105 con T_m impulsada por la concentración. El ligador 103 o los cebadores 105 también pueden unirse a adaptadores universales (no se muestran). Como alternativa a las técnicas de PCR de emulsión descritas anteriormente, la amplificación de cebador enlazado puede lograrse utilizando múltiples etapas de PCR lineal. Durante la PCR lineal, se preparan dos copias de la plantilla genómica 107. Como se muestra en la Figura 1B, el complejo 109 comprende el enlazador 103, los cebadores 105 y las copias idénticas de la plantilla de ácido nucleico 107. Como se muestra en la Figura 1C, se utiliza una segunda etapa de PCR lineal utilizando un enlazador diferente 116 y adaptadores 118 para crear los sentidos 114 opuestos a las plantillas de ácido nucleico 107. Los complejos 109 y 119 se someten a etapas adicionales de amplificación, tales como la PCR universal, para crear múltiples amplicones de ambos sentidos (el sentido y antisentido). Véase para las Figuras 1D y 1E.

En la Figura 3 se muestra un complejo de ejemplo. Como se muestra en la Figura 3, un complejo 301 contiene un ligador 301 unido a dos cebadores 303. El complejo también comprende cebadores de lectura de secuencia 305 y adaptadores 307 para enlazarse al ácido nucleico diana 309. El complejo también comprende adaptadores complementarios 311 y cebadores 313. En una realización preferida, los cebadores 313 son cebadores P7 y los cebadores 303 son cebadores P5. Debe apreciarse que cualquier combinación, orientación o configuración de los adaptadores, cebadores y ácidos nucleicos diana puede organizarse. También debe apreciarse que los complejos pueden incluir códigos de barras. La Figura 3 es un ejemplo y no una realización limitante.

Los complejos de la invención pueden unirse a diversos soportes sólidos tales como microperlas, perlas, paredes de canales, microchips, etc.

La secuenciación de los fragmentos enlazados puede llevarse a cabo por cualquier método conocido en la técnica. La presente invención tiene aplicaciones en diversas plataformas de secuenciación, que incluyen los secuenciadores del genoma de Roche/454 Life Sciences (Margulies et al. (2005) *Nature*, 437:376-380; patente US-6.274.320; 6.258.568; 6.210.891), el sistema SOLiD de Life Technologies Applied Biosystems (Grand Island, NY), el sistema HELISCOPE de Helicos Biosciences (Cambridge, MA) (véase, p. ej., publicación US- 2007/0070349), y los secuenciadores de iones de Life Technologies Ion Torrent, Ion Torrent Systems, Inc. (Guilford, CT).

La secuenciación puede llevarse a cabo mediante métodos donde cada base se determina secuencialmente. Las técnicas de secuenciación de ADN incluyen reacciones clásicas de secuenciación de dideoxi (método de Sanger) mediante la utilización de terminadores marcados o cebadores y separación de gel en la losa o capilar, secuenciación mediante síntesis mediante la utilización de nucleótidos marcados terminados de manera reversible, pirosecuenciación, secuenciación 454, hibridación específica de alelo a una genoteca de sondas oligonucleotídicas marcadas, secuenciación por síntesis mediante la utilización de hibridación específica de alelos con una genoteca de clones marcados que se sigue por el ligamiento, monitoreo en tiempo real y secuenciación SOLiD. La secuenciación de moléculas separadas se ha demostrado más recientemente mediante reacciones de extensión secuenciales o únicas utilizando polimerasas o ligasas, así como mediante hibridaciones diferenciales simples o secuenciales con genotecas de sondas.

Debe apreciarse que el enlazador también puede unirse a adaptadores, cebadores o moléculas de unión. El enlazador puede unirse a estas especies en cualquier orientación o disposición. La molécula de enlace puede unirse directamente a un adaptador o cebador y enlazarse indirectamente a los fragmentos de ácido nucleico. La molécula

de enlace puede eliminarse antes o después de la amplificación. Alternativamente, la molécula de enlace puede permanecer en el complejo. La molécula de enlace puede eliminarse antes de la secuenciación, o la molécula de enlace puede permanecer en el complejo durante la secuenciación.

5 Una técnica de secuenciación que puede utilizarse en los métodos de la invención proporcionada incluye, por ejemplo, la secuenciación de una sola molécula verdadera (tSMS) Helicos (Harris T. D. et al. (2008) Science 320:106-109). En la técnica de tSMS, una muestra de ADN se escinde en cadenas de aproximadamente 100 a 200 nucleótidos, y se añade una secuencia de poliA al extremo 3' de cada cadena de ADN. Cada cadena se marca mediante la adición de un nucleótido de adenosina marcado con fluorescencia. A continuación, las cadenas de ADN se hibridan con una celda de flujo, que contiene millones de sitios de captura oligo-T y que se inmovilizan en la superficie de la celda de flujo. Las plantillas pueden tener una densidad de aproximadamente 100 millones de plantillas/cm². A continuación, la celda de flujo se carga en un instrumento, p. ej., un secuenciador HeliScope, y un láser ilumina la superficie de la celda de flujo, revelando la posición de cada plantilla. Una cámara CCD puede correlacionar la posición de las plantillas en la superficie de la celda de flujo. El marcador fluorescente plantilla se escinde y se lava. La reacción de secuenciación comienza mediante la introducción de una ADN polimerasa y un nucleótido marcado por fluorescencia. El ácido nucleico oligo-T sirve como cebador. La polimerasa incorpora los nucleótidos marcados al cebador de una manera dirigida por plantilla. La polimerasa y los nucleótidos no incorporados se eliminan. Las plantillas que tienen la incorporación dirigida del nucleótido marcado por fluorescencia se detectan mediante imágenes de la superficie de la celda de flujo. Después de la formación de imágenes, una etapa de escisión elimina el marcador fluorescente, y el proceso se repite con otros nucleótidos marcados por fluorescencia hasta que se logre la longitud de lectura deseada. La información de secuencia se recoge con cada etapa de adición de nucleótidos. Con la presente invención, los fragmentos enlazados pueden identificarse en tándem. La descripción adicional de tSMS se muestra, por ejemplo, en Lapidus y col. (patente US-7.169.560), Lapidus y col. (solicitud de patente US-2009/0191565), Quake y col. (patente US-6.818.395), Harris (patente US-7.282.337), Quake y col. (solicitud de patente US-2002/0164629), and Braslavsky, y col., PNAS (USA), 100: 3960-3964 (2003).

25 Otro ejemplo de una técnica de secuenciación de ADN que puede utilizarse en los métodos de la invención proporcionada es la secuenciación 454 Roche) (Margulies, M et al. 2005, Nature, 437, 376-380). La secuenciación 454 implica dos etapas. En la primera etapa, el ADN se somete a cizallamiento en fragmentos de aproximadamente 300-800 pares de bases, y los fragmentos son romos. Los adaptadores de oligonucleótido se ligan a continuación a los extremos de los fragmentos. Los adaptadores sirven como cebadores para la amplificación y secuenciación de los fragmentos. Los fragmentos pueden unirse a perlas de captura de ADN, p. ej., perlas recubiertas con estreptavidina utilizando, p. ej., adaptador B, que contiene una etiqueta de biotina en 5'. Utilizando los métodos descritos, se capturan fragmentos unidos como se describió anteriormente sobre las perlas. Los fragmentos unidos a las perlas se amplifican por PCR dentro de gotas de una emulsión de aceite-agua. El resultado es múltiples copias de fragmentos de ADN amplificados clonalmente en cada perla. En la segunda etapa, las perlas se capturan en pocillos (de tamaño picolitro). La pirosecuenciación se realiza en cada fragmento de ADN en paralelo. La adición de uno o más nucleótidos genera una señal de luz que se registra mediante una cámara CCD en un instrumento de secuenciación. La intensidad de la señal es proporcional al número de nucleótidos incorporados. La pirosecuenciación hace uso de pirofosfato (PPi) que se libera tras la adición de nucleótidos. La PPi se convierte en ATP mediante la ATP sulfurilasa en presencia de adenosina fosfosulfato en 5'. La luciferasa utiliza ATP para convertir luciferina en oxiluciferina, y esta reacción genera luz que se detecta y analiza.

45 Otro ejemplo de una técnica de secuenciación de ADN que se puede utilizar en los métodos de la invención proporcionada es la secuenciación Ion Torrent (solicitudes de patente US-2009/0026082, 2009/0127589, 2010/0035252, 2010/0137143, 2010/0188073, 2010/0197507, 2010/0282617, 2010/0300559), 2010/0300895, 2010/0301398, y 2010/0304982). En la secuenciación Ion Torrent, el ADN se corta en fragmentos de aproximadamente 300-800 pares de bases, y los fragmentos son romos. Los adaptadores de oligonucleótido se ligan a continuación a los extremos de los fragmentos. Los adaptadores sirven como cebadores para la amplificación y secuenciación de los fragmentos. Los fragmentos pueden unirse a una superficie y se fijan a una resolución de manera que los fragmentos sean resolubles individualmente. Utilizando los métodos descritos, los fragmentos unidos se fijan a la superficie. La adición de uno o más nucleótidos libera un protón (H⁺), cuya señal es detectada y registrada en un instrumento de secuenciación. La intensidad de la señal es proporcional al número de nucleótidos incorporados.

La Figura 11 ilustra la preparación de perlas para secuenciación Ion Torrent utilizando PCR de emulsión. Al igual que con los métodos de preparación de plantillas enlazadas descritos anteriormente, un fragmento 701 de muestra de ácido nucleico, tal como ADN libre de células que tiene una región de interés 724, se liga 703 con adaptadores que incluyen cebadores de PCR 727 lineales en un extremo y cebadores de PCR 728 universales en el otro para formar una molécula de plantilla 730. A continuación, se crea una emulsión de 713 que comprende una única molécula plantilla 730, cebadores específicos del gen múltiplex 723, y una pluralidad de cebadores enlazados entre sí sobre una perla 729 en una gota 726. A continuación, la plantilla se amplifica 715 en la gota utilizando métodos de PCR conocidos para crear una perla con numerosas copias de la molécula de plantilla 730, que ahora puede albergar errores introducidos por PCR 721. A continuación, la emulsión se rompe y los fragmentos enlazados a perlas 731 se cargan 717 en el pocillo 732 de secuenciación de la celda de flujo para la secuenciación Ion Torrent. La mayoría de las perlas se perderán durante la carga de la celda de flujo solo con una pequeña fracción del fragmento que contiene las perlas 731 haciéndola en un pocillo de la celda de flujo 732.

65 La Figura 12 ilustra las ventajas de identificación de errores de la secuenciación Ion Torrent utilizando perlas sembradas a partir de una sola molécula de plantilla, donde las bases mezcladas en lecturas de un solo pocillo son

indicativas de un error de PCR mientras las lecturas de consenso que pueden diferir de otros pocillos o de genomas de referencia pueden indicar una variante verdadera en el ácido nucleico de muestra.

5 Se describen métodos de secuenciación de ácidos nucleicos amplificados generados por amplificación en fase sólida. Por lo tanto, también se describe un método de secuenciación de ácido nucleico que comprende amplificar un agrupamiento de plantilla de ácido nucleico utilizando amplificación en fase sólida y llevando a cabo una reacción de secuenciación de ácido nucleico para determinar la secuencia de la totalidad o una parte de al menos una cadena de ácido nucleico amplificada producida en la reacción de amplificación en fase sólida. El punto de inicio para la reacción de secuenciación puede proporcionarse mediante la hibridación de un cebador de secuenciación a un producto de una reacción de amplificación en fase sólida. En este sentido, uno o ambos adaptadores añadidos durante la formación de la genoteca de plantillas pueden incluir una secuencia de nucleótidos que permite la hibridación de un cebador de secuenciación a productos amplificados derivados de genomas completos o amplificación en fase sólida de la genoteca de plantillas.

15 Los productos de reacciones de amplificación en fase sólida en donde ambos cebadores de amplificación directa e inversa se inmovilizan covalentemente en la superficie sólida son las denominadas estructuras puenteadas formadas por hibridación de pares de cadenas de polinucleótidos inmovilizados y cadenas complementarias inmovilizadas, unidas ambas cadenas al soporte sólido en el extremo 5'. Las matrices compuestas por tales estructuras puenteadas proporcionan plantillas ineficientes para técnicas típicas de secuenciación de ácidos nucleicos, ya que la hibridación de un cebador de secuenciación convencional a una de las cadenas inmovilizadas no se ve favorecida en comparación con la hibridación de esta cadena a su cadena complementaria inmovilizada en condiciones estándar para la hibridación.

20 Para proporcionar plantillas más adecuadas para la secuenciación de ácidos nucleicos, puede ser ventajoso eliminar o desplazar sustancialmente toda o al menos una parte de una de las cadenas inmovilizadas en la estructura con puente para generar una plantilla que sea al menos parcialmente monocatenaria. Por lo tanto, la parte de la plantilla que es monocatenaria estará disponible para la hibridación con un cebador de secuenciación. El proceso de eliminación de todo o una porción de una cadena inmovilizada en una estructura de ácido nucleico bicatenario "puente" puede denominarse en la presente memoria como linealización, y se describe con más detalle en la publicación de US 2009/0118128.

25 Las estructuras de plantilla puenteadas pueden linealizarse mediante escisión de una o ambas cadenas con una endonucleasa de restricción o por escisión de una cadena con una endonucleasa de corte. Otros métodos de escisión pueden utilizarse como una alternativa a las enzimas de restricción o las enzimas de corte, que incluyen, entre otras cosas, la escisión química (p. ej., la escisión de un enlace diol con peryodato), la escisión de sitios abásicos por escisión con endonucleasa (por ejemplo "USER", como se suministra por NEB, número de pieza M55055), o por exposición al calor o álcali, la escisión de los ribonucleótidos incorporados en los productos de amplificación que se componen de desoxirribonucleótidos, escisión fotoquímica o escisión de un enlazador peptídico

30 Después de la etapa de escisión, independientemente del método utilizado para la escisión, el producto de la reacción de escisión puede someterse a condiciones de desnaturalización para eliminar la(s) porción(es) de la(s) cadena(s) escindida(s) que no estén unidas al soporte sólido. Las condiciones de desnaturalización adecuadas, por ejemplo, solución de hidróxido de sodio, solución de formamida o calor, serán evidentes para el lector experto con referencia a protocolos estándar de biología molecular (Sambrook et al., supra; Ausubel et al. supra). La desnaturalización da como resultado la producción de una plantilla de secuenciación que es parcial o sustancialmente monocatenaria. Una reacción de secuenciación puede iniciarse a continuación mediante hibridación de un cebador de secuenciación a la porción monocatenaria de la plantilla. Por lo tanto, la invención abarca métodos en los que la reacción de secuenciación de ácido nucleico comprende hibridar un cebador de secuenciación con una región monocatenaria de un producto de amplificación linealizado, incorporar secuencialmente uno o más nucleótidos en una cadena de polinucleótido complementaria a la región de cadena de patrón amplificada a secuenciar, identificar la base presente en uno o más del(de los) nucleótido(s) incorporado(s) y determinar de este modo la secuencia de una región de la cadena de la plantilla.

35 40 45 50 Otro ejemplo de una tecnología de secuenciación que puede utilizarse es la secuenciación Illumina. El flujo de trabajo de secuenciación Illumina se basa en tres etapas: las genotecas se preparan a partir de prácticamente cualquier muestra de ácido nucleico, se amplifican para producir agrupamientos clonales y se secuencian utilizando síntesis masivamente paralela. La secuenciación Illumina se basa en la amplificación del ADN en una superficie sólida mediante la utilización de la PCR de retroceso y cebadores anclados. El ADN genómico se fragmenta y se añaden adaptadores a los extremos 5' y 3' de los fragmentos. Los fragmentos de ADN que se unen a la superficie de los canales de la celda de flujo se extienden y se amplifican en puente. Utilizando los métodos descritos, los fragmentos enlazados se unen a los canales de celda de flujo y se amplifican en puente. El enlazador puede eliminarse antes de la amplificación de puente. Alternativamente, el enlazador permanece unido a los fragmentos durante la amplificación. Los fragmentos se vuelven bicatenarios, y las moléculas bicatenarias se desnaturalizan. Múltiples ciclos de la amplificación en fase sólida seguido de desnaturalización pueden crear varios millones de agrupamientos de aproximadamente 1000 copias de moléculas de ADN monocatenario de la misma plantilla en cada canal de la celda de flujo. Los cebadores, ADN polimerasa y cuatro nucleótidos de terminación reversible, marcados con fluoróforo se utilizan para realizar secuenciación secuencial. Después de la incorporación de nucleótidos, se utiliza un láser para excitar los fluoróforos y se captura una imagen y se registra la identidad de la primera base. Se retiran los terminadores 3' y fluoróforos de cada base incorporada y se repiten las etapas de incorporación, detección e identificación. La secuenciación según esta tecnología se describe en la patente US- 7.960.120; patente US- 7.835.871; patente US- 7.232.656; patente US- 7.598.035; patente US-

6.911.345; patente US- 6.833.246; patente US- 6.828.100; patente US- 6.306.597; patente US- 6.210.891; Publicación US 2011/0009278; Publicación US 2007/0114362; Publicación US 2006/0292611; y la publicación US 2006/0024681.

Los métodos descritos pueden incorporarse en la plataforma de secuenciación Illumina (disponible comercialmente en Illumina, Inc, San Diego, CA). Utilizando los métodos descritos, se preparan genotecas de complejos enlazados que comprenden dos copias idénticas de un fragmento o ambas cadenas de un fragmento dúplex y a continuación se unen al soporte sólido. Los complejos se amplifican para producir agrupamientos clonales y a continuación se secuencian utilizando síntesis masivamente paralela. En este método, cada agrupamiento se siembra con un fragmento. Con los métodos descritos, se utilizan dos fragmentos idénticos o ambas cadenas de un fragmento dúplex para sembrar un agrupamiento. Durante la secuenciación, si existe una falta de acuerdo en una base particular entre los amplicones, se detecta el error.

Los fragmentos unidos pueden unirse a las paredes del canal de celda de flujo. Como se muestra en la Figura. 2, los complejos 109 y 119 están unidos a un soporte sólido 202, tal como una pared del canal de celda de flujo. El complejo 109 puede comprender el sentido y el complejo 119 puede comprender el antisentido. Cada complejo siembra un agrupamiento. Como se muestra en la Figura 2, el complejo 109 siembra el agrupamiento 1 (205) y el complejo 119 siembra el agrupamiento 2 (207).

La Figura 4A representa un ejemplo del complejo 401. El complejo 401 comprende un enlazador 406 y copias idénticas de una plantilla de ácido nucleico. Sin embargo, una copia de la plantilla de ácido nucleico comprende un error 410. El complejo 401 se une al soporte sólido 402 a través de los sitios 405 de unión. En algunos ejemplos, los sitios 405 de unión son oligonucleótidos complementarios (complementarios a oligonucleótidos sobre los complejos) que están unidos covalentemente a la superficie de la celda de flujo. Como se muestra en la Figura 4B, el complejo 401 se extiende y se amplifica en puente para crear copias 418. Este proceso se repite, y como se muestra en la Figura. 4C, se forma un agrupamiento 450 sobre el soporte sólido 402. A partir de este proceso, el agrupamiento es una mezcla de oligonucleótidos derivados de cada mitad de un complejo. Aproximadamente la mitad de los oligonucleótidos contienen el error y la otra mitad no.

La Figura 5A muestra dos complejos 502 y 503, donde el complejo 502 contiene un error 510. A continuación, se introducen cebadores, ADN polimerasa y cuatro nucleótidos de terminación reversible, marcados con fluoróforo para realizar secuenciación secuencial. Después de la incorporación de nucleótidos, se utiliza un láser para excitar los fluoróforos y se captura una imagen y se registra la identidad de la primera base. Dado que no hay error en la primera base, ambas bases emiten fluorescencia iguales. Se retiran los terminadores 3' y fluoróforos de cada base incorporada y se repiten las etapas de incorporación, detección e identificación. Las etapas se repiten hasta que se alcanza la base que contiene el error. En esta base, las bases no emiten fluorescencia lo mismo. Las bases generarían una fluorescencia diferente. Como se muestra en la Figura 5B, la fluorescencia mixta indicaría que las bases no coinciden. La fluorescencia mixta indicaría un error, y la base se notificaría como desconocida, o N. Véase la Figura 5C.

La Figura 16 ilustra un método de la invención basado en gotas para crear ácidos nucleicos dúplex enlazados a partir de las cadenas sentido y antisentido de un fragmento de ácido nucleico. Como se muestra, puede obtenerse un ADN libre de células bicatenario (ADNlc) que tiene una variante rara en ambas cadenas. La plantilla bicatenaria puede añadirse a una emulsión con uno o más cebadores directos específicos de genes (p. ej., la emulsión puede contener cebadores directos e inversos multiplexados específicos para más de un gen o parte de un gen), uno o más cebadores inversos específicos de genes, un cebador unido universal. La emulsión puede someterse a PCR de emulsión para crear productos dúplex enlazados. A continuación, la emulsión puede romperse y digerirse la plantilla sin enlazar. Los productos dúplex enlazados restantes pueden secuenciarse a continuación. Debido a que el producto bicatenario entra en la gota, con cebadores específicos de genes directos e inversos, puede obtenerse información de la secuencia dúplex. Los productos enlazados de la PCR en emulsión contienen ambos sentidos de la plantilla al menos un 50 % de las veces, lo que reduce la tasa media de error. Como se muestra en la Figura 16, se introduce un error de PCR en el producto dúplex durante la PCR de emulsión pero, debido a que el error de PCR solo está presente en una cadena y la variante real está presente en ambos, los dos pueden diferenciarse fácilmente entre sí durante la secuenciación.

Las Figuras 17 y 18 muestran cebadores enlazados universales ilustrativos y cebadores específicos de genes directos e inversos y métodos para su utilización en amplificación por PCR para crear productos dúplex enlazados. Preferentemente, la longitud del amplicón se mantiene corta para mejorar la sensibilidad. En los ejemplos, la región diana entre cebadores es de aproximadamente 86 pb. Se muestran cebadores específicos de genes adicionales en la Figura 19.

Las Figuras 20 y 21 ilustran métodos de secuenciación que utilizan los productos derivados de las Figuras 17 y 18 respectivamente. Los cebadores enlazados pueden contener dos o más sitios y pueden estar hechos de PEG, Trapapap unida a ADN biotinilado, perlas recubiertas de ADN, nanopartículas recubiertas con ADN, perlas enlazadas a ADN a perlas basadas en gel (p. ej., acrilamida). Las perlas pueden ser poliestireno, látex, magnético, sílice, ferromagnéticos o materiales similares. La unión puede ser por métodos convencionales y preferiblemente mediante una combinación de grupos amino y carboxilo.

Otros métodos pueden incluir estrategias de identificación dúplex para moléculas dúplex unidas a gotas formadas. Como se señaló, los métodos de la invención basados en gotas pueden dar como resultado al menos un 50 % de tasa de formación de fragmentos dúplex enlazados (moléculas enlazadas que contienen representaciones de cada lado del dúplex de ADN),

por lo que la identificación de esos productos se vuelve importante para omitir los datos de los productos que no son dúplex y recoger los beneficios crecientes de precisión de los productos dúplex. Los métodos de identificación dúplex pueden incluir, por ejemplo, un enfoque de PCR de dos etapas que utiliza dos conjuntos de cebadores con diferentes temperaturas de hibridación donde varios ciclos iniciales se realizan a baja temperatura con cebadores de código de barras específicos de genes para amplificar e identificar cada sentido del dúplex, mientras se añade una cola universal para ciclos posteriores. El número de ciclos de código de barras se limita para evitar etiquetar cada sentido del dúplex con múltiples códigos de barras. Los ciclos posteriores pueden realizarse a alta temperatura a través de cebadores universales porque los cebadores de código de barras no pueden unirse en esas condiciones. Los productos dúplex pueden identificarse por la presencia de sus códigos de barras específicos de detección durante el análisis de secuenciación.

Las Figuras 29A-D ilustran métodos de identificación dúplex. En el ejemplo ilustrativo, puede añadirse lo siguiente a la gota: un cebador de enlace; un cebador directo universal y un cebador inverso universal, cada uno con una alta T_m (T_m puede aumentarse utilizando LNA); un cebador específico de gen directo de código de barras y un cebador específico de gen inverso con código de barras, cada uno con una T_m más baja y a una concentración más baja que el cebador directo universal; y la plantilla dúplex. La PCR de emulsión puede ejecutarse con un primer ciclo que tiene una temperatura de hibridación baja para permitir que los cebadores con código de barras se unan a la plantilla seguido de un segundo ciclo de temperatura de hibridación bajo para producir los productos mostrados en la Figura 29B. Un tercer ciclo de temperatura de hibridación bajo permite el primer ciclo de unión a cebador universal. En este ciclo, los cebadores con código de barras todavía se unirán [A + B] para formar más o los productos C y D, y también pueden unirse a los productos C y D para formar más productos E y F.

Después del tercer ciclo, los productos que se muestran en la Figura 29C pueden estar presentes en la emulsión, que a continuación pueden someterse a un cuarto ciclo de temperatura de hibridación bajo para permitir un segundo ciclo de unión a cebador universal. Al final del ciclo 4, las moléculas con las colas universales directas e inversas completas pueden obtenerse como se muestra en la Figura 29C. La temperatura de hibridación puede aumentarse para ciclos posteriores. Puede haber algunos productos de tipo I y J que tengan diferentes códigos de barras (p. ej., tienen colas universales completas en el lado directo o inverso). Solo pueden amplificarse linealmente a una temperatura de hibridación más alta.

Los ciclos de PCR posteriores (5+) pueden tener una temperatura de hibridación aumentada solo permitiendo la unión de cebadores universales a amplicones con una cola universal completa como se muestra en la Figura 29D. Los últimos ciclos pueden estar a una temperatura de hibridación baja para permitir el enlace de las cadenas amplificadas a través de una parte de la cola universal directa. Alternativamente, puede utilizarse un cebador enlazado más largo con la cola universal directa completa que permita el enlace a temperaturas de hibridación más altas pero es más difícil de sintetizar y puede ser menos eficiente para el enlace. El sentido superior o inferior de enlace se produce al azar, por lo que el 50 % de las moléculas enlazadas mediante la utilización de este cebador unido deben tener 1 de cada una (información dúplex). Los cebadores enlazados con más de 2 sitios, por ejemplo 100 sitios en una nanopartícula, en promedio contienen información dúplex casi 100 % del tiempo.

En ciertas realizaciones, pueden crearse moléculas dúplex enlazadas sin la utilización de PCR de emulsión. En realizaciones que no se basan en gotas, puede utilizarse un único ciclo de amplificación para crear una molécula dúplex enlazada que tenga tanto las cadenas sentido como antisentido del fragmento original. La molécula dúplex enlazada puede cargarse a continuación directamente en una celda de flujo para su secuenciación, evitando así los sesgos de secuencia o longitud inducidos por la amplificación o (p. ej., en la secuenciación del genoma completo), así como evitar los errores introducidos por la amplificación y las pérdidas de ácido nucleico debidas a una eficiencia de carga deficiente. Por ejemplo, cuando la eficiencia de carga de un secuenciador puede definirse como: (número de lecturas de salida)/(número de moléculas de entrada capaces de formar lecturas), la eficiencia de carga para Illumina MiSeq es <0,1 %, y es similar para otros instrumentos de Illumina. Esto se debe en gran medida a pérdidas fluidicas, ya que más de 600 ul de muestra se cargan en el secuenciador, mientras que solo ~7 ul se retiene dentro de la celda de flujo en la unión, lo que da como resultado grandes pérdidas de material de partida. Los métodos de carga directa que no se basan en gotas descritos en la presente memoria reúnen estas ineficiencias. Los métodos de la invención pueden utilizar un flujo de trabajo simplificado que crea una molécula dúplex con un ciclo de PCR. Las moléculas dúplex pueden utilizarse para sembrar uno solo agrupamiento y proporcionar lecturas de secuenciación de alta precisión. Al cargar la celda de flujo directamente y a continuación secuenciar, se minimizan las pérdidas de ADN a través de la carga.

La carga directa, los métodos no de gota de la invención tienen aplicaciones que incluyen secuenciación del genoma completo donde está presente una pequeña masa de ADN, pero se desea una alta precisión, tal como biopsia de tejido, aspirados con aguja o extracciones de sangre de pequeño volumen. Las aplicaciones adicionales pueden incluir aquellas en las que el ADN se degrada o daña, tal como en muestras fijadas con formalina, embebidas en parafina (FFPE).

Las Figuras 22A, 22B, 23A y 23B muestran métodos de formación dúplex enlazados que no están basados en gotas según ciertas realizaciones de la invención. Uno (Figura 22) o dos (Figura 23) adaptadores de enlace se ligan sobre la plantilla genómica bicatenaria y a continuación se extienden utilizando una polimerasa de desplazamiento de cadena para crear la molécula dúplex enlazada. El dúplex enlazado puede cargarse directamente en una celda de flujo para secuenciación. En dos aplicaciones de adaptador de enlace, tal como se ilustra en la Figura 23, los fragmentos enlazados pueden formarse en dos orientaciones (es decir, fragmentos enlazados que tengan el enlazador en un extremo y fragmentos enlazados que tengan el enlazador en el extremo opuesto). Como se muestra

en las Figuras 22B y 22B, la ligación puede dar como resultado aproximadamente el 50 % del producto dúplex enlazado deseado donde otros productos no deseados no formarán agrupamientos.

Las Figuras 24A y 24B ilustran etapas de un método de secuenciación de carga directa utilizando moléculas dúplex enlazadas. En el método de ejemplo de las Figuras 24A y 24B, una celda de flujo se inicializa con reactivos. A continuación se desnaturaliza una genoteca unida A volumen pequeño y el volumen completo cargado en la celda de flujo inicializada. A continuación, los puertos de celda de flujo se sellan y la plantilla tal como se crea en los métodos ilustrados en las Figuras 22 y 23, se une a la celda de flujo. El ADN en la celda de flujo se extiende y a continuación la celda de flujo se carga en el instrumento de secuenciación de celda de flujo. La unión a la celda de flujo de ejemplo se ilustra en la Figura 25, que incluye las etapas de captura de celdas de flujo, extensión, lavado de plantilla enlazada, amplificación de puente y secuenciación. La unión para la otra cadena de sentido de la plantilla dúplex enlazada es análoga a la ilustrada en la Figura 25.

La Figura 26 ilustra un protocolo de siembra fuera de línea de ejemplo en comparación con un protocolo predeterminado. Las etapas del protocolo de siembra fuera de línea pueden incluir realizar las siguientes etapas en el banco a temperatura ambiente: al ras con LDR x 5, al ras con PR2 x 5, al ras con HT1, carga TMP y puertos de sellado con cinta de PCR, donde el enjuague significa llenar la celda de flujo con el reactivo especificado, esperando aproximadamente 10 segundos y a continuación vaciar la celda de flujo. Después de sellar los puertos con cinta de PCR, la celda de flujo se incuba en un baño de perlas a 75 grados Celsius durante 10 minutos, seguido de incubación a 40 grados Celsius durante 10 minutos. Al devolver la celda de flujo al banco a temperatura ambiente, se retira el precinto y se lava la celda de flujo con PR2 a 40 grados Celsius 5 veces, se lava con AMS 1 2 veces, se lava con AMS 1 con una incubación de dos minutos a 40 grados Celsius 3 veces, se llena con AMS1 y se transfiere a un instrumento MiSeq (disponible comercialmente en Illumina, Inc, San Diego, CA) para la secuenciación. Las etapas adicionales en el protocolo de preparación pueden incluir tomar la celda de flujo fuera de su carcasa de plástico, precortar cinta de PCR para orificios de sellado y proteger la celda de flujo de arañazos del baño de perlas, con cinta de PCR o cinta de extracción en ambos lados.

Para la carga directa, así como otras aplicaciones donde el rendimiento de la carga de la celda de flujo y el rendimiento de captura de la diana son importantes, puede ser beneficioso combinar la carga de las celdas de flujo con secuenciación dirigida, para minimizar la pérdida. Dicha combinación simplifica adicionalmente el flujo de trabajo eliminando una etapa adicional. Si bien existen métodos para la captura de objetivos en la celda de flujo, sufren de al menos dos lados inferiores. En primer lugar, no son capaces de secuenciar la región que se captura en la celda de flujo. Para fragmentos cortos tales como ADN libre de células, esto puede tener una gran pérdida de señal. En segundo lugar, son incapaces de capturar moléculas dúplex enlazadas, como se describe en la invención, para su secuenciación. Por consiguiente, se describen métodos de captura de diana basada en celdas de flujo de moléculas dúplex. La celda de flujo puede contener un sentido de oligonucleótidos (oligos) con regiones diana, mientras que los otros sentidos están capilarizados y no disponibles inmediatamente para la unión. Véase la Figura 27. Después de que un sentido de moléculas enlazadas es capturado en la celda de flujo, los otros oligos de celdas de flujo se activan para capturar el otro sentido de los fragmentos enlazados (p. ej., mediante la utilización de una digestión de uracilo, digestión enzimática o luz). La plantilla puede a continuación extenderse y la generación de agrupamientos puede continuar tan normal. En ciertas realizaciones, un conjunto de oligos puede ser complementario a la cadena sentido o antisentido del ácido nucleico dúplex mientras que el otro conjunto es complementario a un adaptador universal que se ha unido tanto a las cadenas sentido como antisentido y los oligos adaptadores universales pueden ser capilarizados para impedir la unión en una etapa de exposición inicial.

Las Figuras 28A-E ilustran las etapas de un método ejemplar para la captura de diana basada en celdas de flujo de moléculas dúplex. La Figura 28A muestra una etapa de captura de diana ilustrativa donde una molécula enlazada se carga en una celda de flujo, ya sea directamente o mediante métodos convencionales. La Figura 28B muestra una etapa ilustrativa de unión de la plantilla a la celda de flujo donde la molécula enlazada se une a una región de captura complementaria, y el otro sentido de oligos de celdas de flujo se libera para unirse a ambos extremos libres del fragmento unido. La Figura 28C muestra una etapa de desplazamiento de cadena a modo de ejemplo donde la polimerasa de desplazamiento de cadena se utiliza para extender ambos fragmentos para crear un agrupamiento de siembra doble. La plantilla enlazada puede a continuación desnaturalizarse y eliminarse de la celda de flujo como se muestra en la Figura 28D. La amplificación de puente puede producirse entonces como algo normal, pero con dos moléculas que siembran el agrupamiento como se muestra en la Figura 28E.

Las técnicas de carga directa de la invención pueden utilizarse en aplicaciones de secuenciación de genoma completo sin etapas de captura de la diana de celda de flujo con uno o dos adaptadores de enlace. En aplicaciones de secuenciación dirigida, después del ligamiento con uno o dos adaptadores enlazados, puede utilizarse una técnica de captura de dianas basada en tubos que se optimiza para el rendimiento (p. ej., que tenga un rechazo deficiente fuera de la diana pero alto rendimiento). La plantilla dúplex enlazada puede cargarse directamente en la celda de flujo como se ha descrito anteriormente con o sin las etapas de captura de destino descritas en las FIGS. 28A-E. Puede omitirse la etapa de captura de la diana basada en tubos intermedios.

El analizador de genomas de Illumina (detector, comercializado por Illumina) se basa en una lectura paralela basada en fluorescencia de millones de secuencias inmovilizadas que se secuencian iterativamente utilizando química de terminador reversible. En un ejemplo, se hibridan hasta ocho genotecas de ADN con una celda de flujo de ocho carriles. En cada uno de los carriles, las moléculas de genoteca monocatenaria hibridan con oligonucleótidos complementarios

que están unidos covalentemente a la superficie de la celda de flujo. La cadena inversa de cada molécula de genoteca se sintetiza y la molécula enlazada ahora se amplifica adicionalmente en un proceso denominado amplificación de puente. Esto genera agrupamientos que contengan cada uno más de 1000 copias de la molécula de partida. A continuación se elimina selectivamente una cadena, los extremos libres se bloquean posteriormente y se hibrida un cebador de secuenciación en las secuencias adaptador de las moléculas del agrupamiento.

Aunque el sistema de formación de imágenes fluorescentes no es lo suficientemente sensible como para detectar la señal de una sola molécula de plantilla, el detector es sensible para detectar la señal de cada agrupamiento. En este ejemplo, se analizan las señales de numerosos agrupamientos. Se espera que cada agrupamiento emita fluorescencia en un valor, por ejemplo, aproximado a una de las cuatro bases. Si el agrupamiento no emite fluorescencia a un valor aproximado a una de las cuatro bases, entonces se determina que existe un error en ese locus.

Después de la secuenciación, se analizan las imágenes y se extraen las intensidades para cada agrupamiento. La llamada de bases de Illumina, Bustard, tiene que manejar dos efectos de los cuatro valores de intensidad extraídos para cada ciclo y agrupamiento; primero, una fuerte correlación de las intensidades A y C, así como de las intensidades G y T debido a espectros de emisión similares de los fluoróforos y separación limitada por los filtros utilizados; y segundo, dependiendo de la señal para un ciclo específico en la señal de los ciclos antes y después, conocidos como ajuste de fase y preajuste de fase, respectivamente. El ajuste y el preajuste de fase son causados por una eliminación incompleta de los terminadores 3' y fluoróforos, las secuencias en el agrupamiento faltan un ciclo de incorporación, así como por la incorporación de nucleótidos sin terminadores 3' eficaces. El ajuste y el preajuste de fase hacen que las intensidades extraídas para un ciclo específico consistan en la señal del ciclo actual, así como el ruido de los ciclos anteriores y siguientes.

Otro ejemplo de una tecnología de secuenciación que puede utilizarse en los métodos de la invención proporcionada incluye la tecnología de una sola molécula, en tiempo real (SMRT) de Pacific Biosciences. En SMRT, cada una de las cuatro bases de ADN está unida a uno de cuatro colorantes fluorescentes diferentes. Estos tintes están reforzados con fosfato. Una única ADN polimerasa se inmoviliza con una sola molécula de ADN monocatenaria plantilla en la parte inferior de una guía de ondas en modo cero (ZMW). Un ZMW es una estructura de confinamiento que permite la observación de la incorporación de un solo nucleótido por ADN polimerasa contra el fondo de nucleótidos fluorescentes que se difunden rápidamente en un valor de ZMW (en microsegundos). Lleva varios milisegundos para incorporar un nucleótido en una cadena de crecimiento. Durante este tiempo, la etiqueta fluorescente se excita y produce una señal fluorescente, y la etiqueta fluorescente se escinde. La detección de la fluorescencia correspondiente del colorante indica qué base se incorporó. El proceso se repite. Utilizando estos métodos descritos, el proceso se repite en tándem, con dos fragmentos que se analizan.

Otro ejemplo de una técnica de secuenciación que puede utilizarse en los métodos de la invención proporcionada es la secuenciación de nanoporos (Soni G V and Meller A. (2007) Clin Chem 53: 1996-2001). Un nanoporo es un pequeño orificio, del orden de 1 nanómetro de diámetro. La inmersión de un nanoporo en un fluido conductor y la aplicación de un potencial a través de ella da como resultado una ligera corriente eléctrica debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño del nanoporo. Como una molécula de ADN pasa a través de un nanoporo, cada nucleótido en la molécula de ADN obstruye el nanoporo en un grado diferente. Por lo tanto, el cambio en la corriente que pasa a través del nanoporo como la molécula de ADN pasa a través del nanoporo representa una lectura de la secuencia de ADN. Utilizando los métodos descritos, dos fragmentos se analizan simultáneamente o secuencialmente, reduciendo la posibilidad de un error.

La presente invención puede utilizarse con tecnología de nanoporos, tal como la secuenciación basada en nanoporos de molécula única mediante síntesis (Nano-SBS). Esta estrategia puede distinguir cuatro bases detectando 4 etiquetas de diferentes tamaños liberadas de nucleótidos modificados con fosfato en 5'. A medida que cada nucleótido se incorpore a la cadena de ADN en crecimiento durante la reacción de la polimerasa, su etiqueta se libera y entra en un nanoporo por orden de liberación. Esto produce una firma de bloqueo de corriente iónica única debido a la estructura química distinta de la etiqueta, determinando así la secuencia de ADN electrónicamente a nivel de molécula única con resolución de base única. Utilizando los métodos descritos, dos fragmentos idénticos o ambas cadenas de un fragmento dúplex pueden analizarse simultáneamente o secuencialmente. Véase Kumar, et al. Scientific Reports, Article number 684, doi: 10.1038/srep00684.

Las funciones descritas anteriormente tales como el análisis o ensamblaje de lectura de secuencia pueden implementarse utilizando sistemas que incluyen software, hardware, firmware, cableado o combinaciones de cualquiera de estos.

Un método de secuenciación que puede utilizarse con la invención se basa en la utilización de nucleótidos modificados que tienen bloques extraíbles en 3', por ejemplo, como se describe en WO04018497, patente US-2007/0166705A1 y patente de US-7.057.026. Una vez que el nucleótido modificado se ha incorporado en la cadena de polinucleótidos de crecimiento complementaria a la región de la plantilla que está secuenciándose, no hay ningún grupo 3'-OH libre disponible para dirigir más extensión de secuencia y, por lo tanto, la polimerasa no puede añadir nucleótidos adicionales. Una vez que se ha determinado la naturaleza de la base incorporada en la cadena de crecimiento, el bloque 3' puede eliminarse para permitir la adición del siguiente nucleótido sucesivo. Al pedir los productos derivados utilizando estos nucleótidos modificados, es posible deducir la secuencia de ADN de la plantilla de ADN. Dichas reacciones pueden realizarse en un solo experimento si cada uno de los nucleótidos modificados tiene una etiqueta diferente unida al mismo, que se sabe que corresponde a la base particular, para facilitar la

discriminación entre las bases añadidas durante cada etapa de incorporación. Alternativamente, puede llevarse a cabo una reacción separada que contenga cada uno de los nucleótidos modificados por separado.

5 Las Figuras 32 y 33 ilustran métodos de secuenciación alternativos que utilizan sistemas y métodos descritos en la presente memoria. Como se muestra en la Figura 32, después de sembrar agrupamientos con múltiples copias de
 10 plantillas y amplificar, los errores pueden diferenciarse de las variantes verdaderas mediante una disminución de la calidad de secuenciación en una sola lectura en la posición donde las bases no sean las mismas. Debido a que todas las cadenas amplificadas en el agrupamiento se leen al mismo tiempo, en la misma dirección, una disminución de la calidad de la señal es la única forma de determinar una llamada de base mixta dentro del agrupamiento. Cuando un
 15 agrupamiento se siembra con una cadena o plantillas sentido y antisentido que tengan diferentes sitios de cebadores de secuenciación, se pueden identificar variantes y errores verdaderos comparando los resultados de dos lecturas de secuenciación diferentes (p. ej., lecturas de cada sentido o lecturas utilizando los dos cebadores de secuenciación diferentes). La Figura 33 ilustra métodos de la invención utilizando dos lecturas de secuenciación separadas para comparar llamadas base de una lectura sentido y antisentido. La secuenciación u otros errores introducidos solo deben verse en una de las lecturas, mientras que las variantes verdaderas deben observarse en ambas lecturas.

Los métodos de captura de dianas enlazadas pueden incluir capturas basadas en soluciones de regiones genómicas de interés para la secuenciación de ADN dirigida. Las Figuras 34 y 35 ilustran métodos ejemplares de captura de dianas basadas en soluciones. Las sondas universales y códigos de barras opcionales (que pueden ser específicos de un sentido) se ligan al ADN extraído. El producto de ADN ligado se desnaturaliza y se une con sondas de captura de dianas enlazadas que comprenden un sitio de cebado universal y sonda universal enlazada a una sonda específica de diana. La captura de dianas se realiza a una temperatura donde las sondas universales no pueden unirse solas a menos que la concentración local sea alta debido a la unión de la sonda diana. A continuación, se utiliza la polimerasa de desplazamiento de cadenas (p. ej., BST, phi29 o SD) para extender las sondas enlazadas a la diana. La sonda diana está bloqueada para la extensión como se indica por el diamante negro en las Figuras 34 y 35, de manera que la extensión solo se produce a lo largo de la sonda universal unida, copiando la cadena de ácido nucleico diana que permanece enlazada a la sonda diana. A continuación, puede utilizarse un número de ciclos de extensión de PCR enlazados para amplificar las secuencias diana. La PCR puede realizarse utilizando cebadores universales correspondientes a los sitios de cebado universales de las sondas de captura de diana enlazadas para amplificar una o ambas cadenas del ácido nucleico diana. Esta etapa de PCR puede realizarse en la misma reacción sin la necesidad de una etapa de limpieza. La secuencia diana amplificada puede secuenciarse como se ha descrito anteriormente. No se requiere ningún espacio entre las sondas de captura enlazadas cuando se utilice en direcciones opuestas, aunque es posible que haya un espacio. Las sondas de captura pueden producirse utilizando enlazadores universales 5' mediante la unión de los enlazadores universales a una sonda de captura prefabricada. Las sondas de captura pueden unirse mediante estreptavidina/biotina u otros medios como se describió anteriormente y el enlazador universal puede extenderse utilizando la sonda de captura como plantilla.

Los métodos de la invención incluyen captura de dianas basada en gotas, opcionalmente utilizando cebadores enlazados universales, para capturar moléculas dúplex. Los métodos basados en gotas representados en la Figura 36 son similares a los ilustrados en la Figura 16 pero utilizan sondas de captura de dianas enlazadas como se ha descrito anteriormente y representado en las Figuras 34-35. Las sondas universales y códigos de barras opcionales (que pueden ser específicos de sentido) se ligan al ADN extraído (p. ej., ADN sin células). Se crea una emulsión como se describió anteriormente utilizando una molécula de plantilla dúplex y sondas de captura de diana que comprenden un sitio de cebado universal y sonda universal enlazada a una sonda específica de diana. Como anteriormente, a captura de la diana se realiza a una temperatura en la que las sondas universales no pueden unirse por sí solas a menos que la concentración local sea elevada debido a la unión de la sonda diana y las sondas de captura estén bloqueadas para no extenderse por sí mismas, pero incluyan un sitio de cebado universal de forma que los cebadores universales y los cebadores universales enlazados incluidos en la emulsión puedan utilizarse para amplificar el ácido nucleico diana y producir una molécula dúplex enlazada que comprenda las cadenas en sentido y antisentido del ácido nucleico diana. Pueden omitirse enlazadores universales para realizar la captura de diana sola. A continuación, la emulsión puede romperse y la plantilla sin enlazar puede digerirse enzimáticamente dejando solo moléculas dúplex enlazadas, a continuación pueden sembrarse agrupamientos o de otro modo secuenciarse como se describió anteriormente.

Las Figuras 37A y B proporcionan detalles adicionales de los métodos de captura de dianas de la invención basados en gotas. La etapa 0 en la Figura 37A muestra que una molécula de plantilla dúplex con sondas universales y códigos de barras opcionales ligados a ella se carga en una gota con cebadores enlazados y universales y sondas de captura de dianas. El ADN plantilla se desnaturaliza en la gota y las sondas de captura diana se unen a las cadenas plantilla desnaturalizadas a una temperatura donde la sonda universal no se unirá sola a menos que la sonda diana también esté unida. El cebador universal a continuación solo se une a las dianas capturadas. La extensión con la polimerasa de desplazamiento de cadena se produce a continuación solo en las dianas capturadas. Pasando a la Figura 37B, se realizan ciclos de extensión (p. ej., 4-6 ciclos) hasta que se agoten las sondas de captura y los cebadores de la diana. Los productos de extensión resultantes se amplifican a continuación utilizando los cebadores enlazados universales para producir moléculas dúplex enlazadas con códigos de barras específicos de cadena. Al igual que con los métodos basados en la solución, no se requiere ningún espacio entre las sondas de captura enlazadas cuando estén en direcciones opuestas. Las sondas de captura enlazadas pueden utilizarse en una o ambas direcciones si se omiten los enlazadores universales para realizar la captura de la diana sola. Las

polimerasas convencionales pueden mezclarse con las polimerasas de desplazamiento de cadena dentro de la gota para llevar a cabo las diversas etapas de extensión y amplificación del método.

5 Ciertos métodos de la invención se refieren a la captura de dianas de moléculas enlazadas. Las copias enlazadas de moléculas tales como las creadas utilizando los métodos descritos anteriormente pueden ser dirigidas y capturadas y convertirse en moléculas enlazadas para su secuenciación. Las Figuras 38-42 ilustran métodos ejemplares de captura de la diana de nanopartículas de moléculas enlazadas. La Figura 38 muestra una nanopartícula que tiene cebadores universales y una cadena que comprende una región diana complementaria a una región de captura de la molécula enlazada a capturar. La Figura 39 ilustra la unión de la región de captura a la región diana. Esta etapa se produce a una temperatura donde se unirán las regiones diana/de captura pero los cebadores universales no se unirán. Las plantillas que no estén unidas pueden lavarse en esta etapa. La temperatura de la reacción puede reducirse para permitir la unión del cebador universal. La Figura 40 muestra la unión de los cebadores universales a sitios de cebadores universales en la molécula enlazada. La Figura 41 muestra la extensión de cebador universal mediante la polimerasa de desplazamiento de cadena para producir copias enlazadas a nanopartículas de la molécula diana que comprende ambas cadenas de la molécula unida original. La Figura 42 muestra una nanopartícula de siembra dual que puede utilizarse para sembrar un agrupamiento en un secuenciador de celdas de flujo como se describe en otra parte de la solicitud.

Pueden incorporarse nucleótidos modificados. Los nucleótidos modificados pueden marcarse (p. ej., marcador fluorescente) para la detección. Cada tipo de nucleótido puede llevar por tanto una etiqueta fluorescente diferente, por ejemplo, como se describe en la publicación US- 2010/0009353. No es necesario que el marcador detectable sea un marcador fluorescente. Puede utilizarse cualquier etiqueta que permita la detección de un nucleótido incorporado. Un método para detectar nucleótidos marcados con fluorescencia comprende la utilización de luz láser de una longitud de onda específica para los nucleótidos marcados, o la utilización de otras fuentes de iluminación adecuadas. La fluorescencia del marcador en el nucleótido puede detectarse mediante una cámara CCD u otros medios de detección adecuados. La instrumentación adecuada para registrar imágenes de matrices agrupadas se describe en WO07123744 y la publ. US- 2010/0111768.

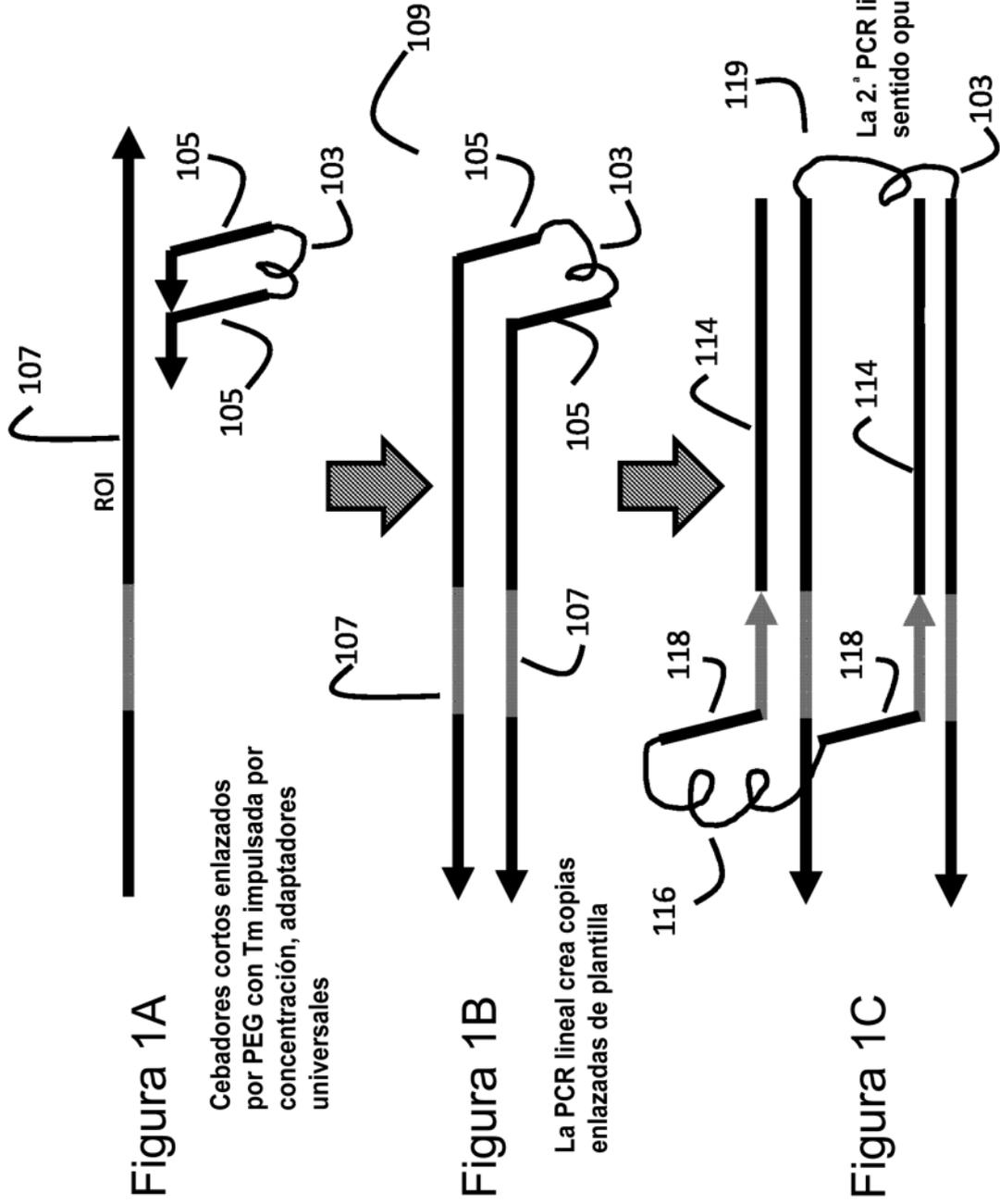
Ejemplo 1: Reducción de errores de secuenciación en el amplicón KRAS utilizando agrupamientos de siembra doble

Los agrupamientos de celdas de flujo se sembraron con moléculas plantilla individuales. Las copias de plantilla individuales fueron de una genoteca de plantillas enlazadas donde solo una de las moléculas plantilla enlazadas estaba unida a la celda de flujo como se muestra en la Figura 31. A continuación, se analizaron los 3000 primeros agrupamientos de siembra individual que se alinearon con el amplicón KRAS en busca de errores de secuenciación con un umbral de calidad aplicado superior a 35. Los agrupamientos de siembra individual dieron como resultado un error medio del 0,13 % con una profundidad media de aproximadamente 3000 como se muestra en la Figura 30. Debido a que la celda de flujo de siembra individual utilizó una genoteca de plantillas enlazadas, los resultados pueden representar una tasa de error más baja que la que se experimentaría utilizando un método estándar de siembra individual con moléculas de plantilla sin enlazar.

A continuación, los agrupamientos de celdas de flujo se sembraron de manera dual utilizando moléculas plantilla enlazadas donde ambas moléculas enlazadas fueron unidas a la celda de flujo para sembrar el agrupamiento. A continuación, se analizaron los 3000 primeros agrupamientos de siembra dual que se alinearon con el amplicón del chr12 se analizaron para detectar errores de secuenciación con el mismo umbral de calidad aplicado mayor que 35 y un filtro de color fluorescente de 0,8 o mayor. Los agrupamientos de siembra dual proporcionaron una reducción multiplicada por 7 de los errores de secuenciación con menos del 3 % de pérdida de bases analizadas respecto a los agrupamientos de siembra individual. La tasa media de error en los agrupamientos de siembra doble fue de 0,02 % con una profundidad media de aproximadamente 2920 como se muestra en la Figura 43.

REIVINDICACIONES

1. Un método para capturar regiones genómicas de interés para secuenciación de ADN dirigida, el método comprende:
- 5 ligar dianas de sonda universal sobre una pluralidad de fragmentos de ácido nucleico dúplex en donde la pluralidad de fragmentos de ácido nucleico dúplex comprende al menos una región genómica de interés,
 10 y en donde la diana de sonda universal es común en toda la pluralidad de fragmentos de ácido nucleico;
 desnaturalizar la pluralidad de fragmentos de ácido nucleico dúplex ligados para crear fragmentos de ácido nucleico monocatenario que comprendan dianas de sonda universales;
 exponer los fragmentos de ácido nucleico monocatenario a una pluralidad de sondas de captura enlazadas que comprenden una sonda diana complementaria a al menos una porción de la región genómica de interés, la sonda diana enlazada en su extremo 5' a 5' de un ácido nucleico que comprende, en un orden de 3' a 5', una sonda universal y un sitio de cebado universal, en donde la etapa de exposición se produce en condiciones que requieren la unión de la sonda diana a la región genómica de interés para permitir la unión de la sonda universal a la diana de la sonda universal, y en donde la sonda diana está bloqueada para la extensión;
 20 extender la sonda universal utilizando una polimerasa de desplazamiento de cadena para producir una copia de la región genómica de interés;
 amplificar la región genómica de interés utilizando amplificación por PCR y cebadores universales complementarios a los sitios de cebado universal; y
 secuenciar la región genómica de interés.
- 25 2. El método de la reivindicación 1, en donde las etapas de desnaturalización, exposición, extensión y amplificación se realizan dentro de una gota de emulsión.
- 30 3. El método de la reivindicación 1, en donde la etapa de ligamiento del sitio de la sonda universal comprende además ligar secuencias de código de barras únicas sobre la pluralidad de fragmentos de ácido nucleico dúplex.
- 35 4. El método de la reivindicación 3, en donde las secuencias de código de barras únicas son específicas de tal manera que las cadenas sentido y antisentido de un fragmento de ácido nucleico de la pluralidad de fragmentos de ácido nucleico son identificables como que se han originado a partir del fragmento de ácido nucleico.
- 40 5. El método de la reivindicación 1, que comprende además unir la sonda diana y la sonda universal juntas utilizando una molécula de unión.
- 45 6. El método de la reivindicación 5, en donde la sonda diana y la sonda universal están biotinizadas y la molécula de enlace comprende estreptavidina.
- 50 7. El método de la reivindicación 1, que comprende además repetir las etapas de exposición y extensión para amplificar la región genómica de interés antes de la etapa de secuenciación.
- 55 8. El método de las reivindicaciones 1 o 7, que comprende además amplificar la región genómica de interés utilizando amplificación por PCR y cebadores universales complementarios a los sitios de cebado universales.
9. El método de la reivindicación 8, en donde los cebadores universales están enlazados de tal manera que la etapa de amplificación produce copias enlazadas de la región genómica de interés.
10. El método de la reivindicación 9, en donde los cebadores universales enlazados son específicos del sentido, de tal manera que la etapa de amplificación produce copias enlazadas de las cadenas sentido y antisentido de la región genómica de interés.
11. El método de la reivindicación 5, en donde la sonda diana y la sonda universal están enlazadas mediante química clic.



Comenzar la PCR universal a lo largo de muchos amplicones, en ambos sentidos

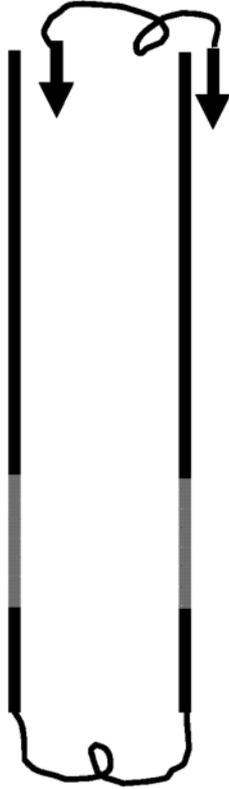


Figura 1D



Moléculas listas para secuenciación

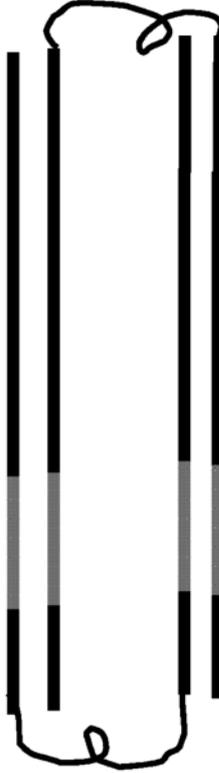


Figura 1E

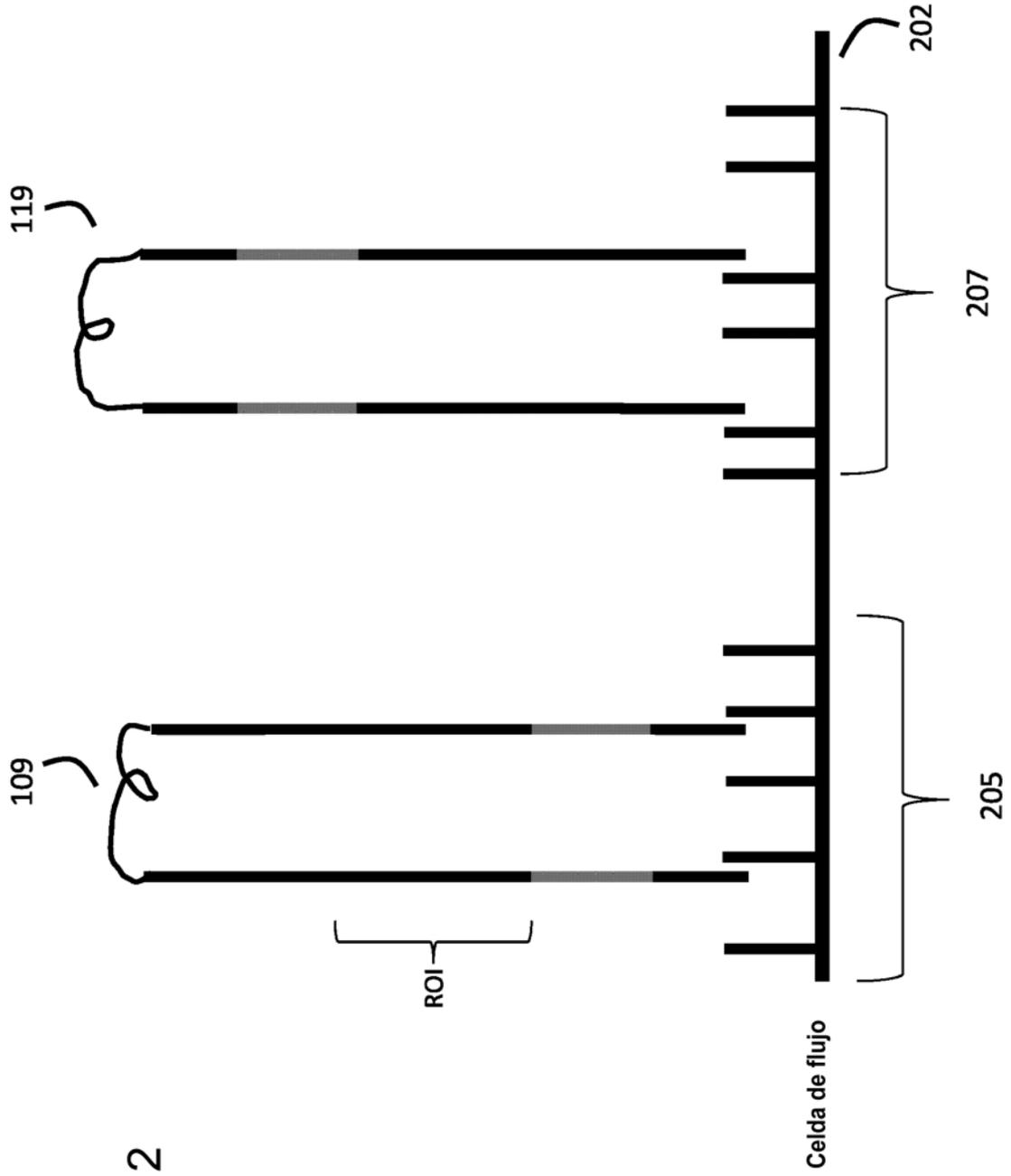
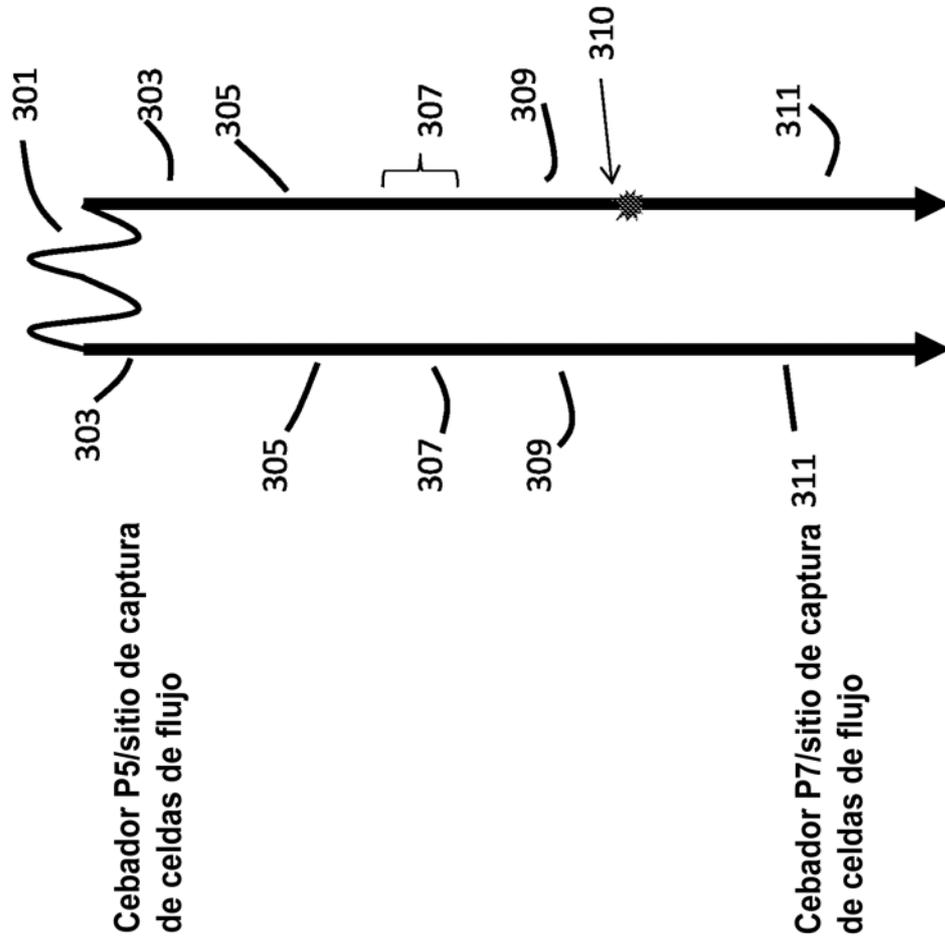


Figura 2

Figura 3



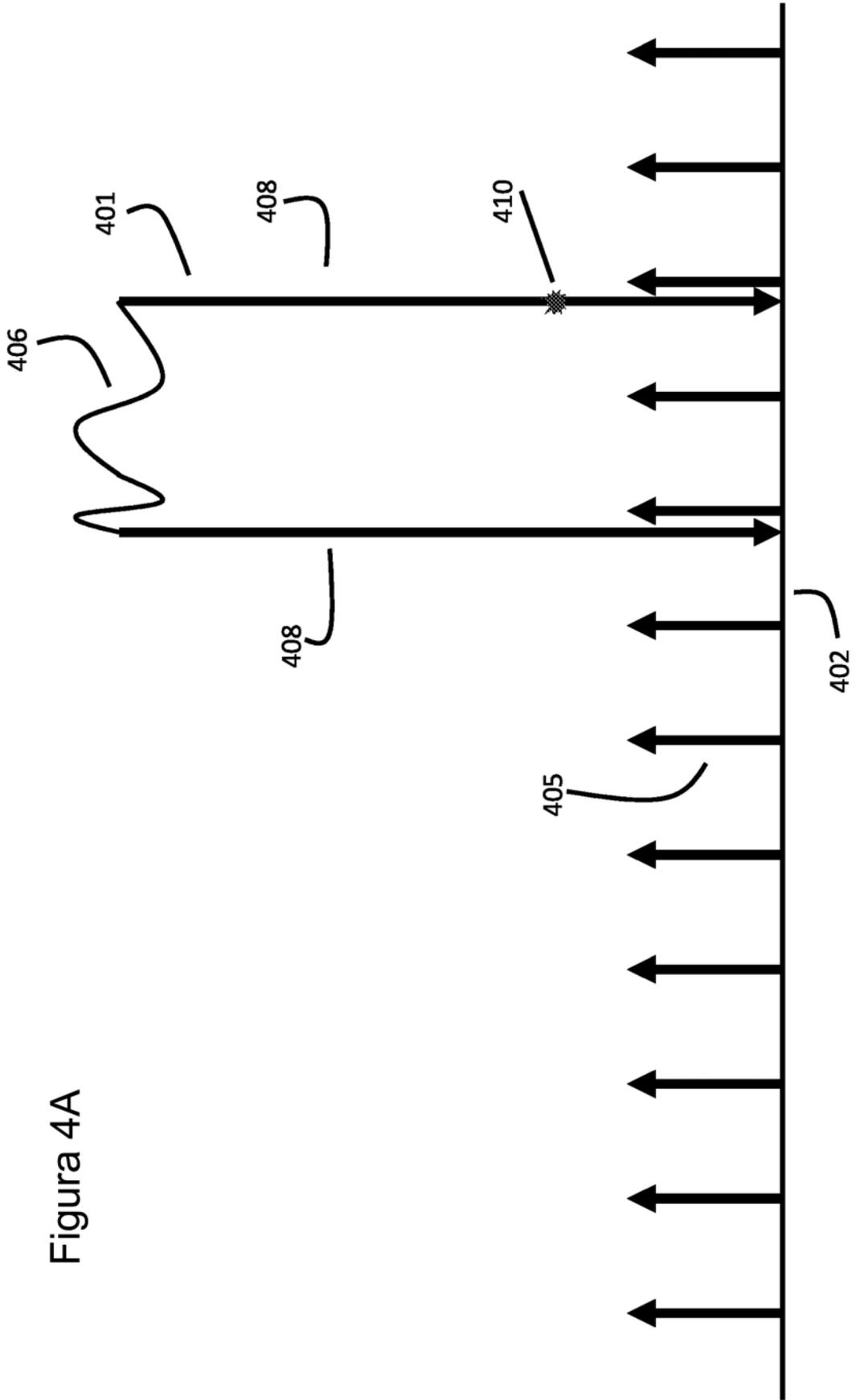
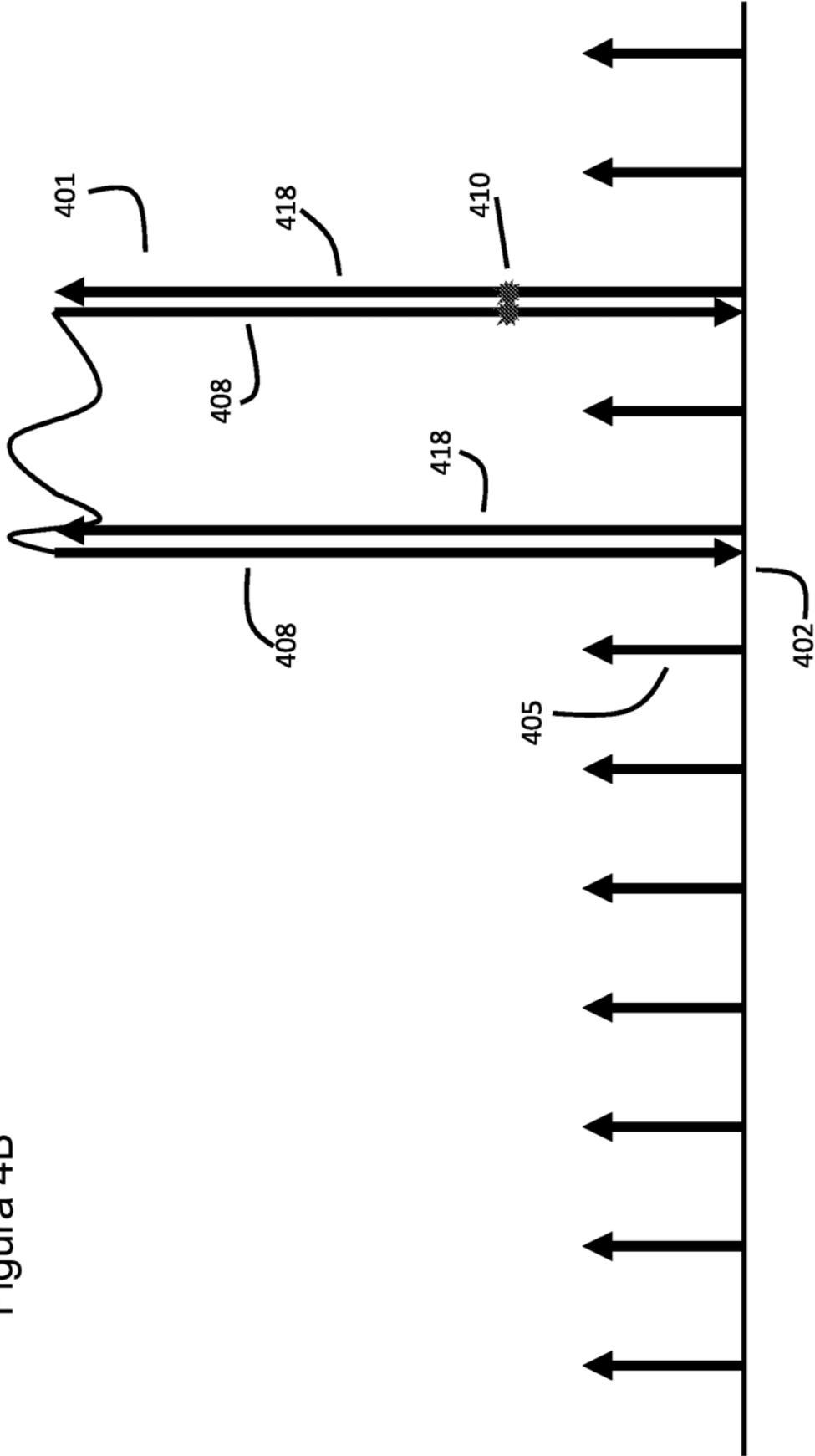


Figura 4A

Figura 4B



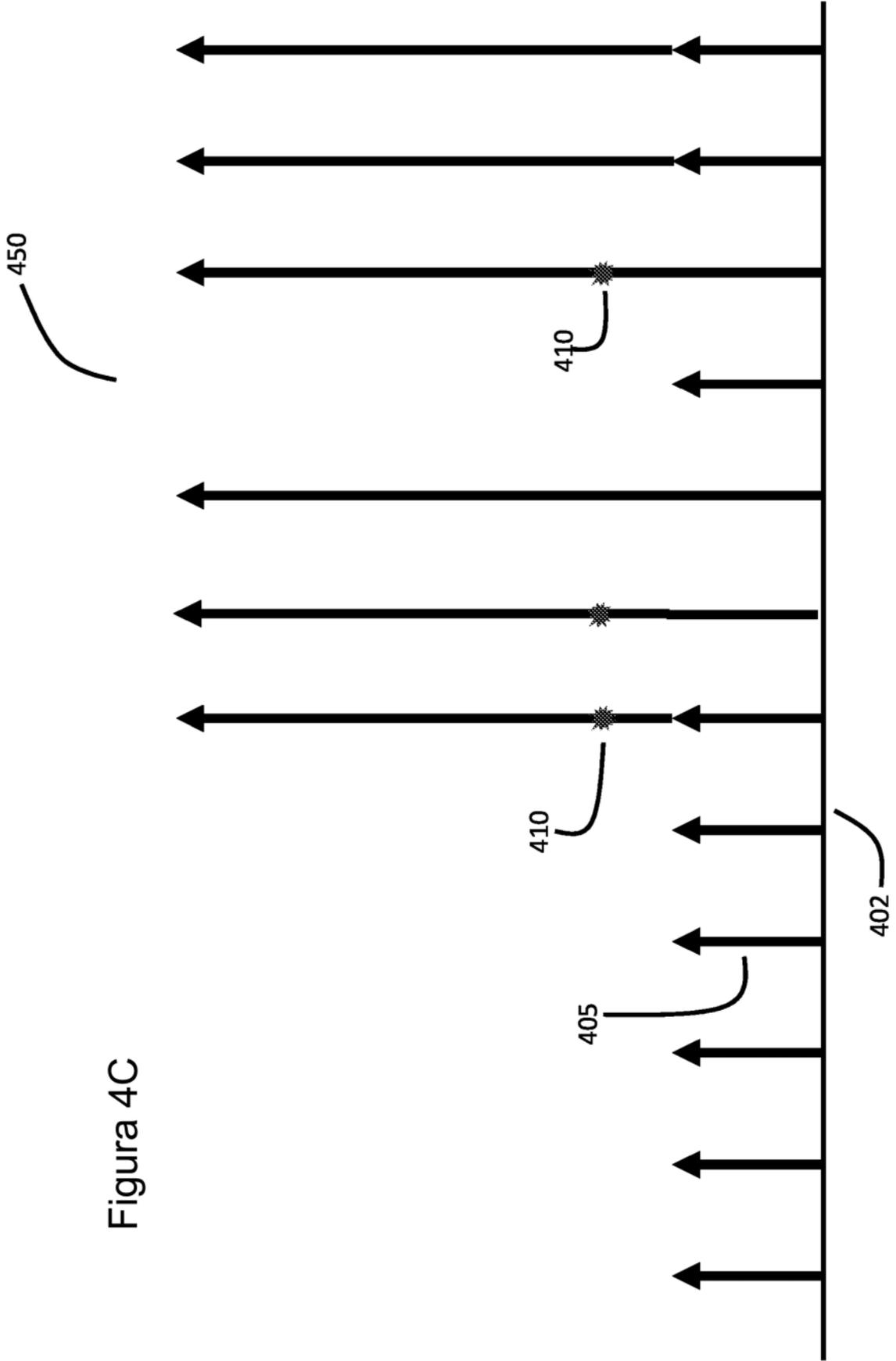


Figura 4C

Figura 5A

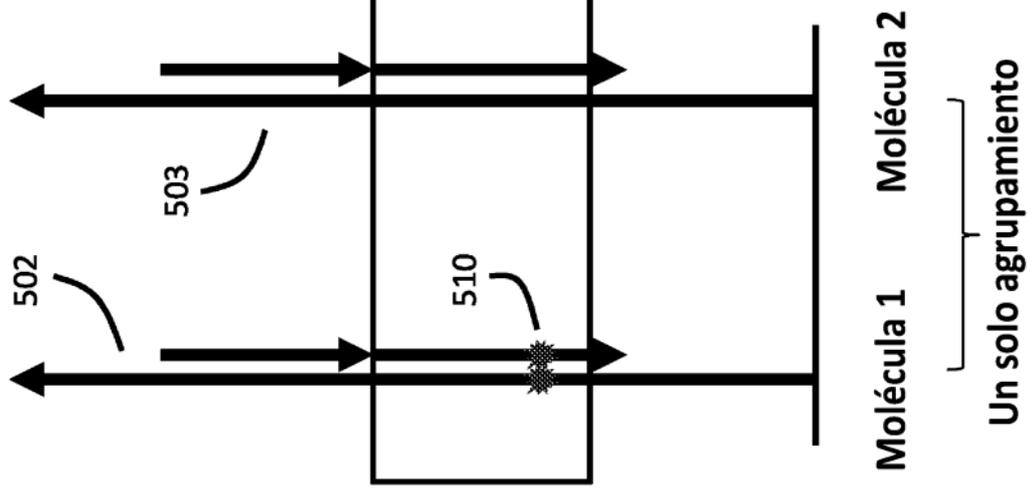
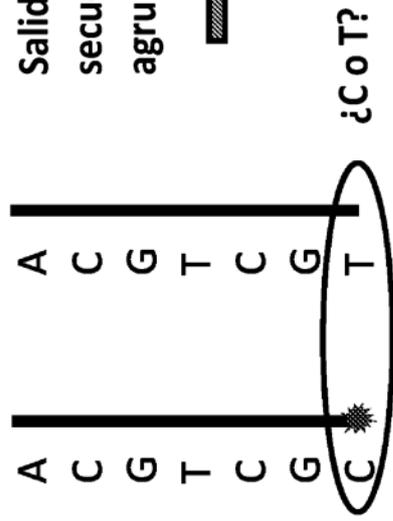


Figura 5B



Salida de
secuenciación del
agrupamiento → ACGTCGN

Estimación de la
calidad de la base

Figura 5C

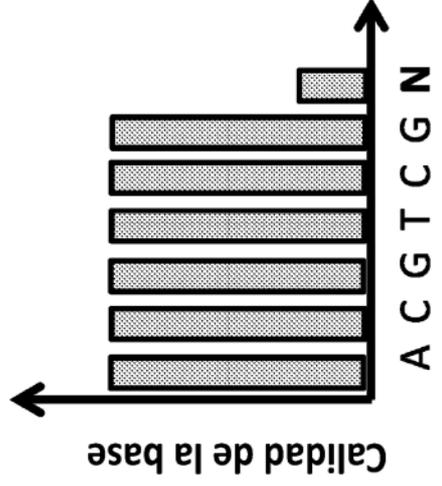
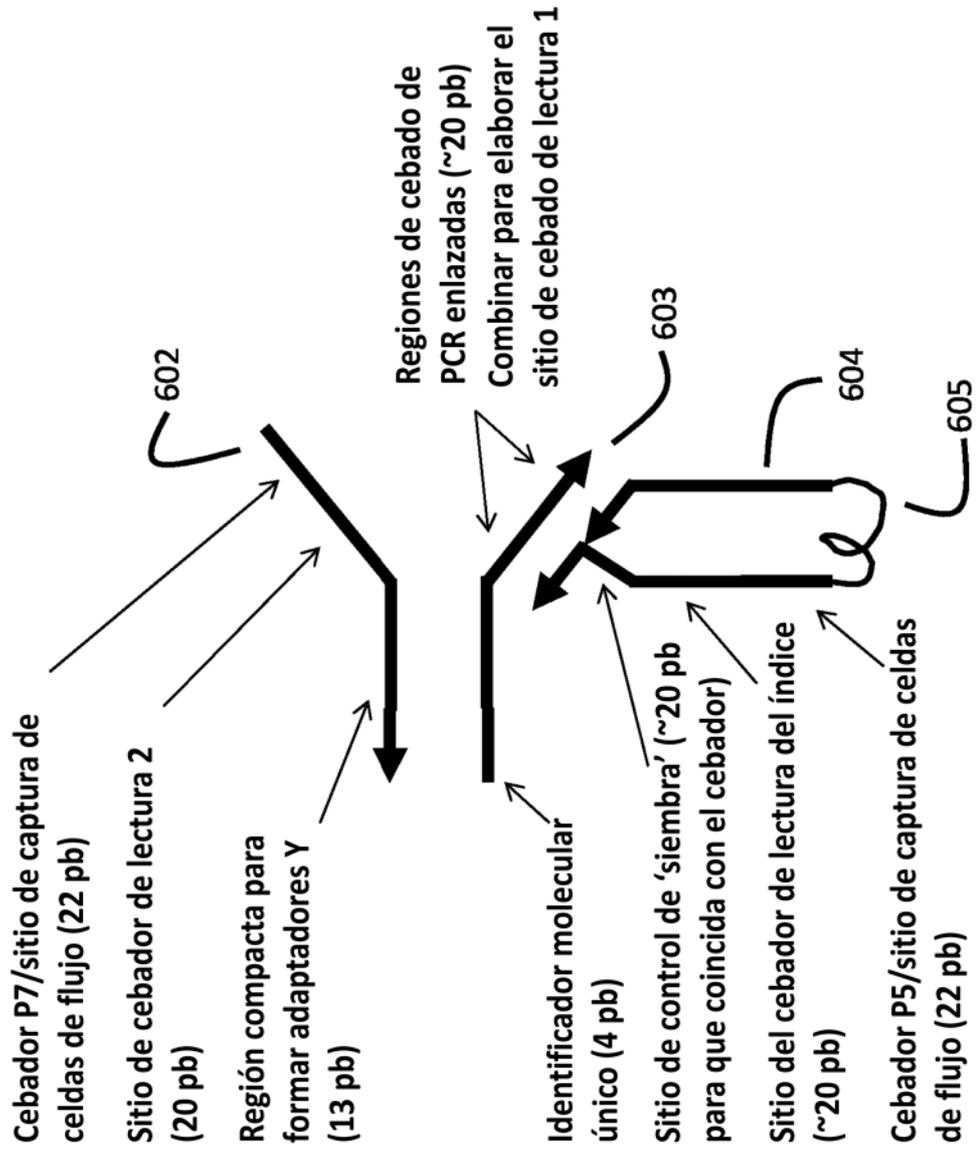


Figura 6



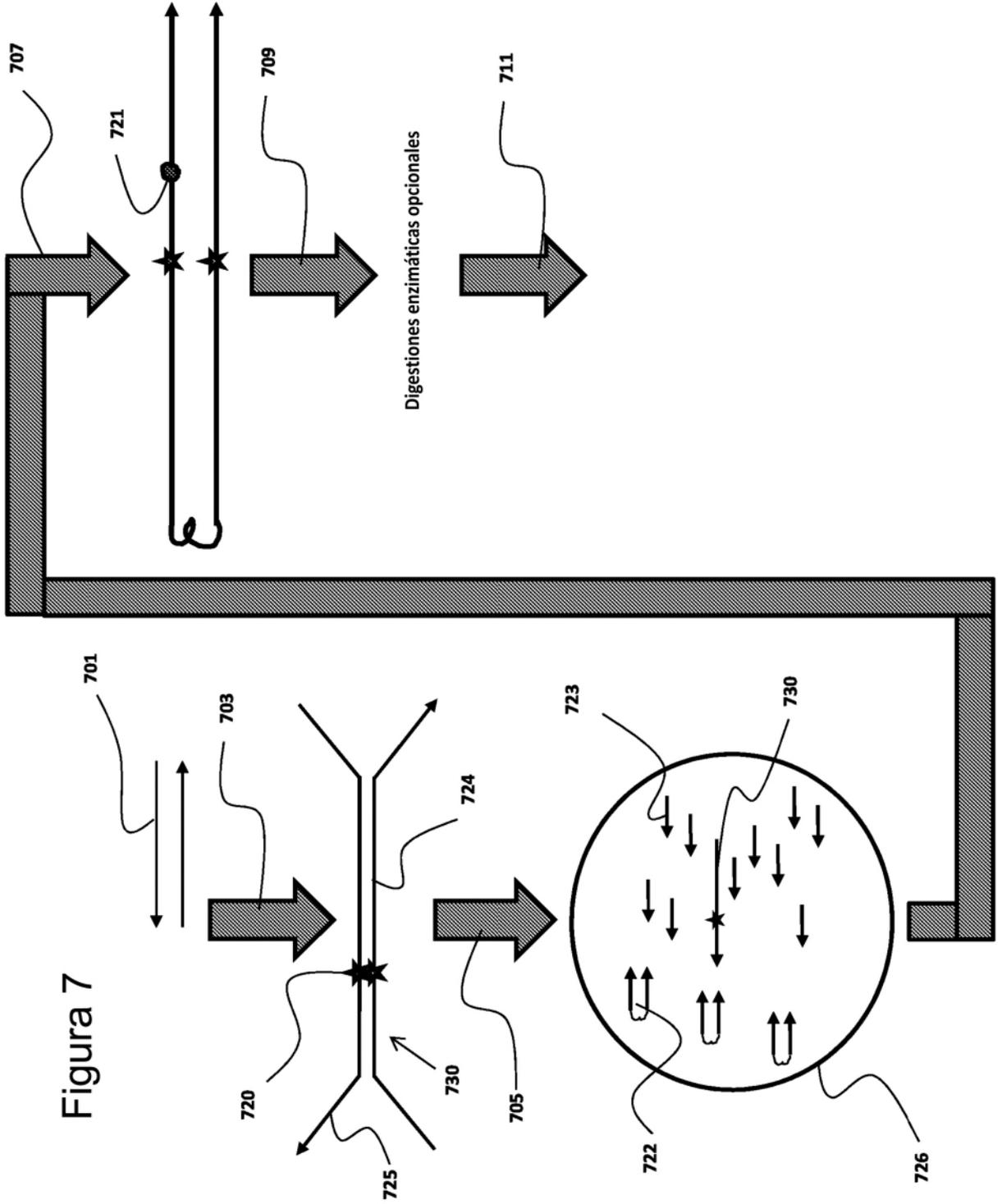


Figura 7

Figura 8

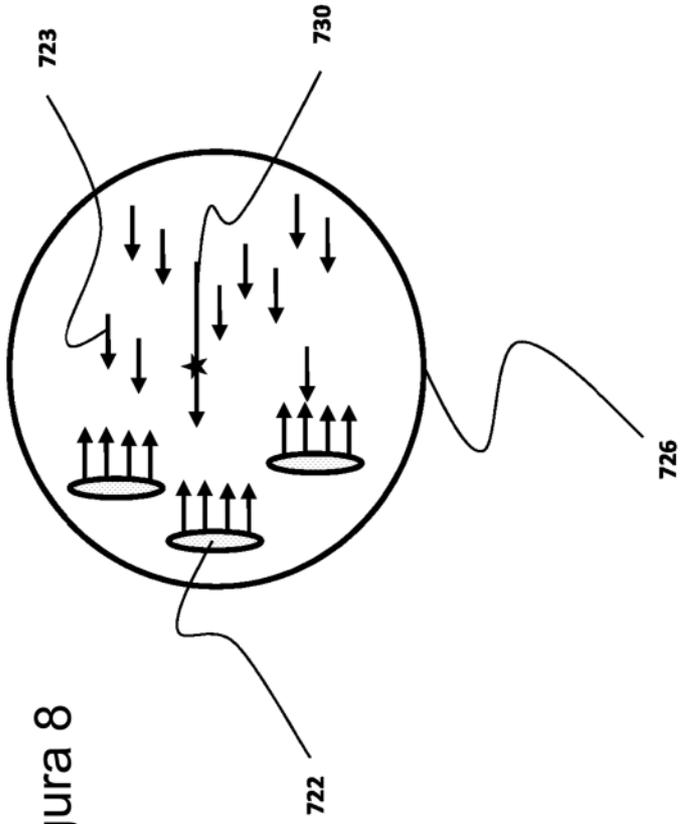


Figura 9

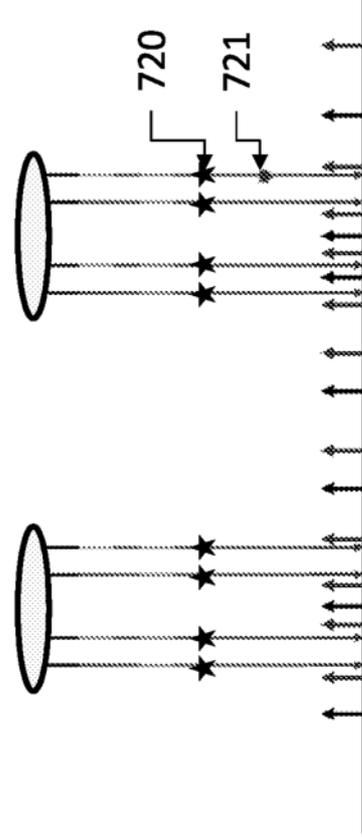


Figura 9

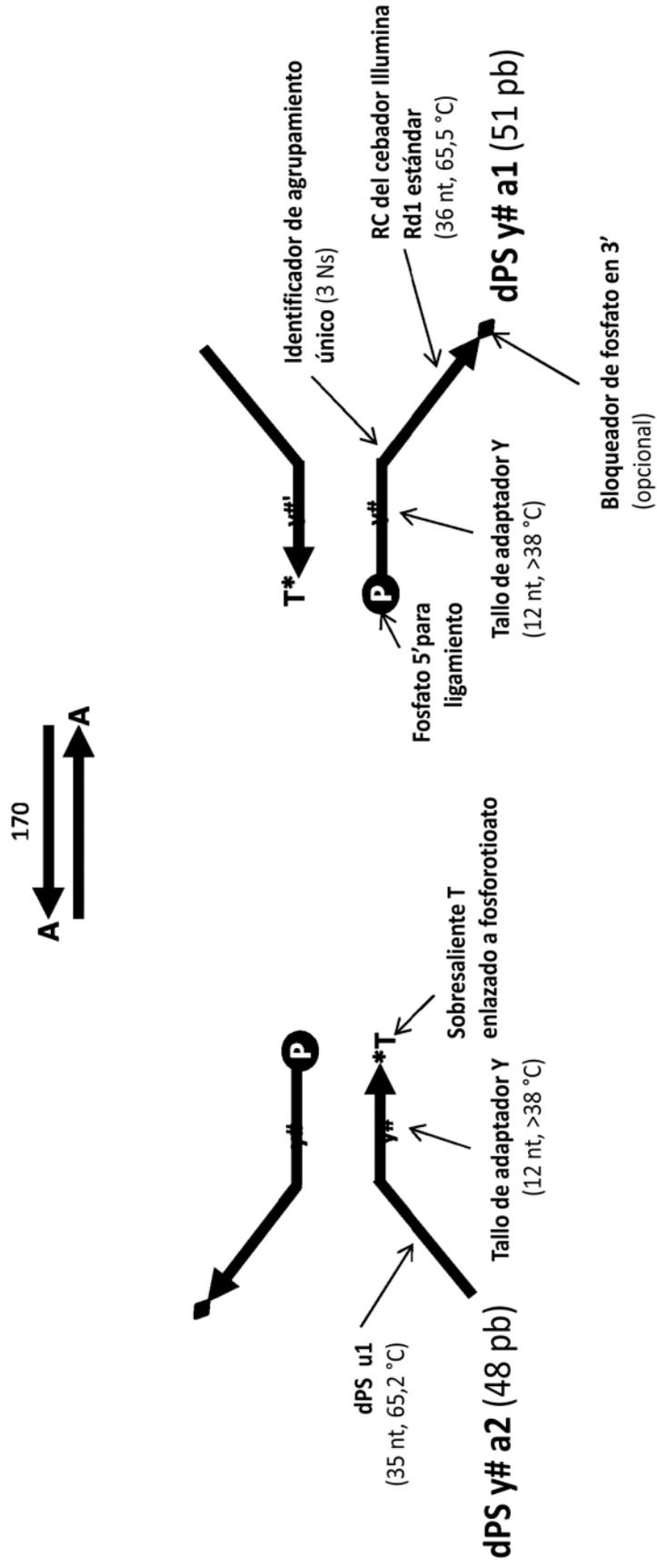
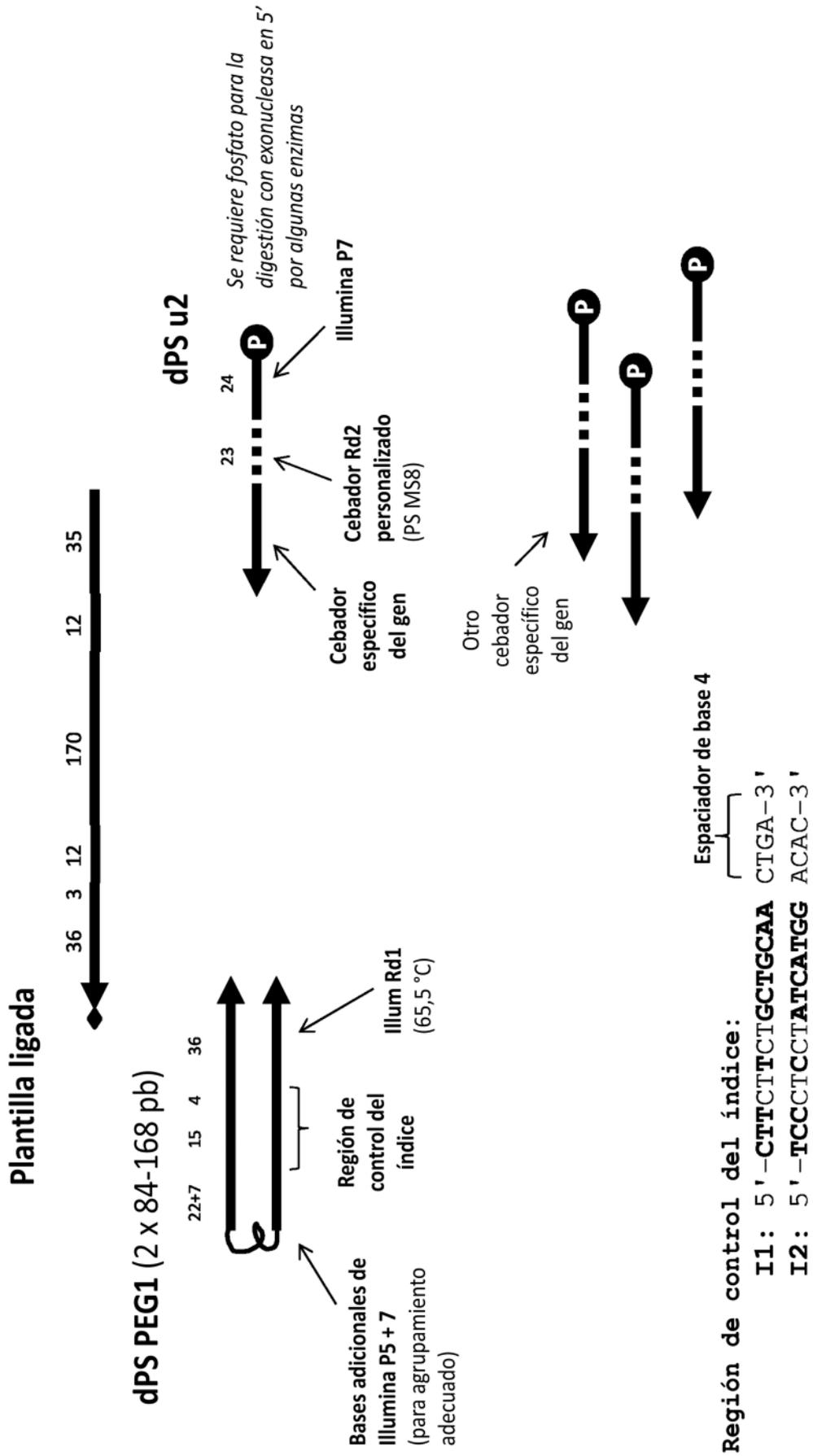


Figura 10



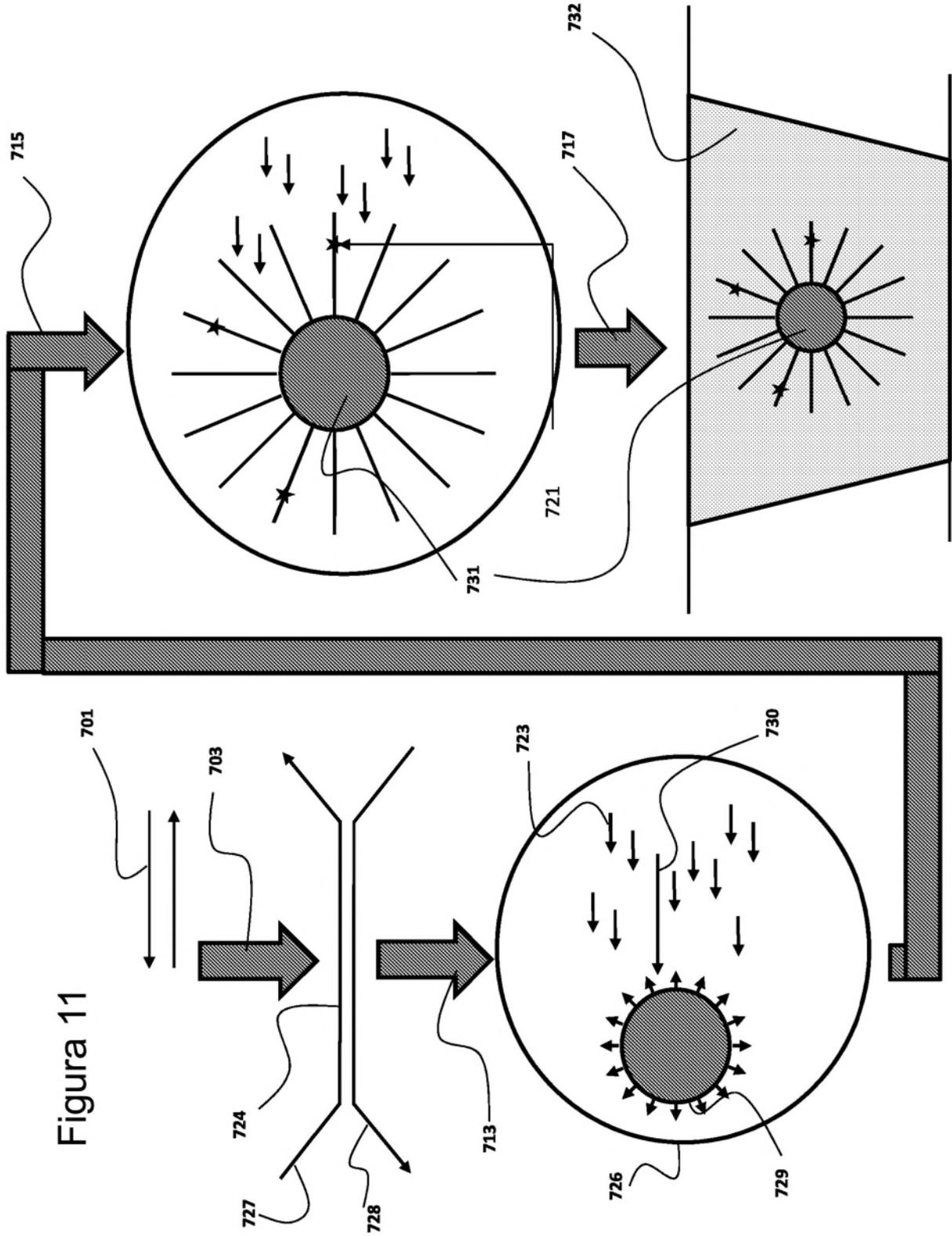
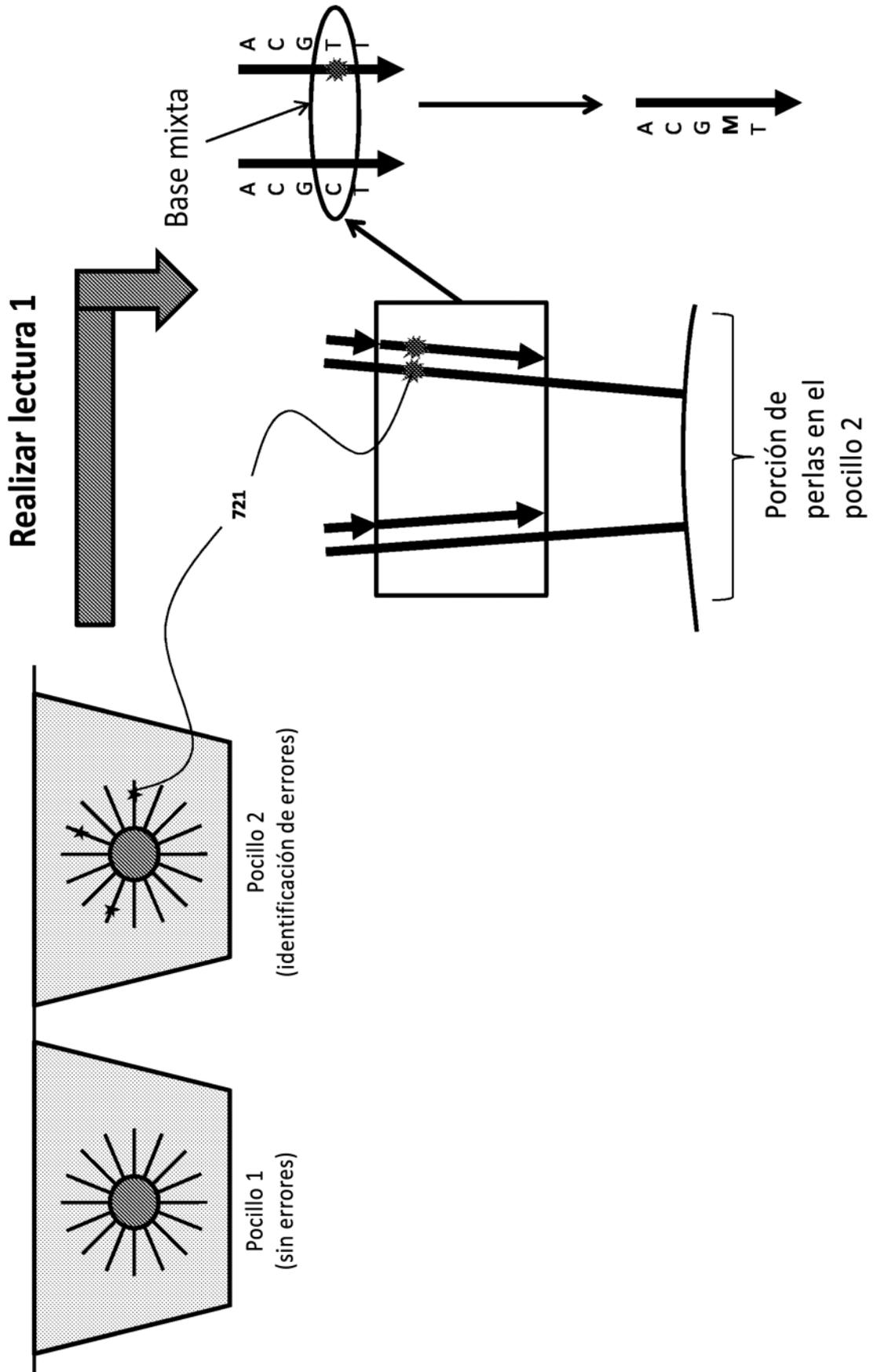


Figura 11

Figura 12



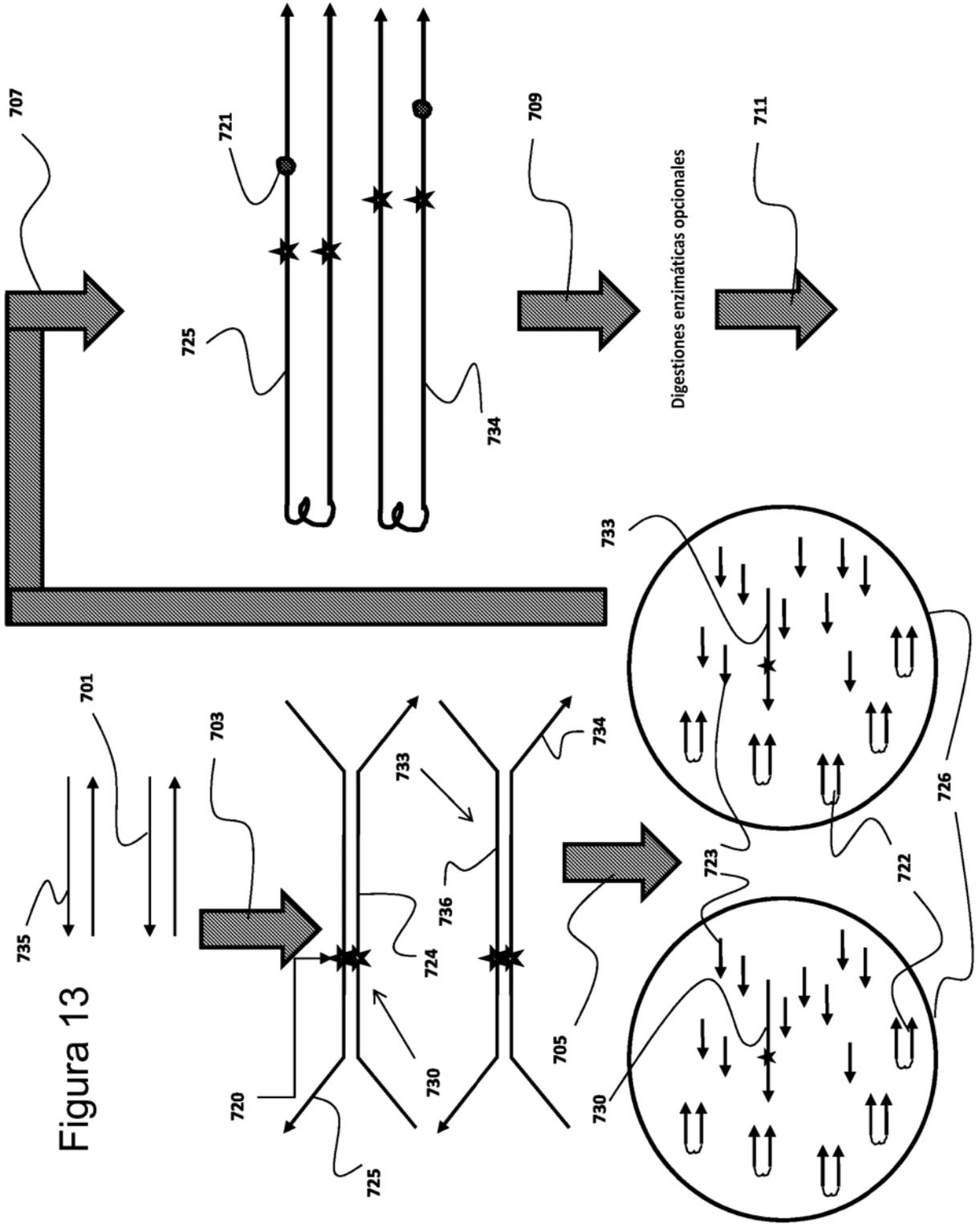


Figura 13

Figura 14

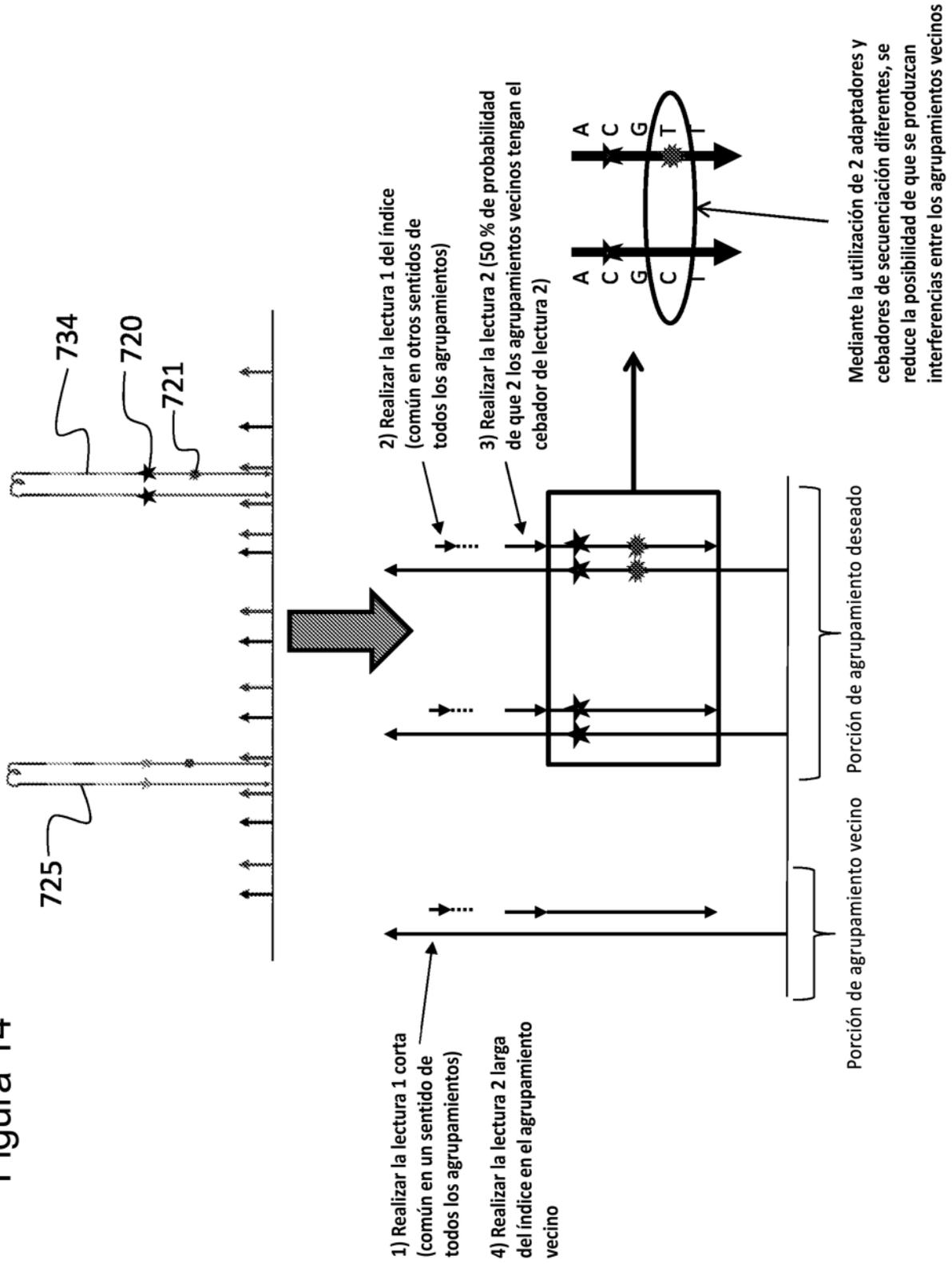
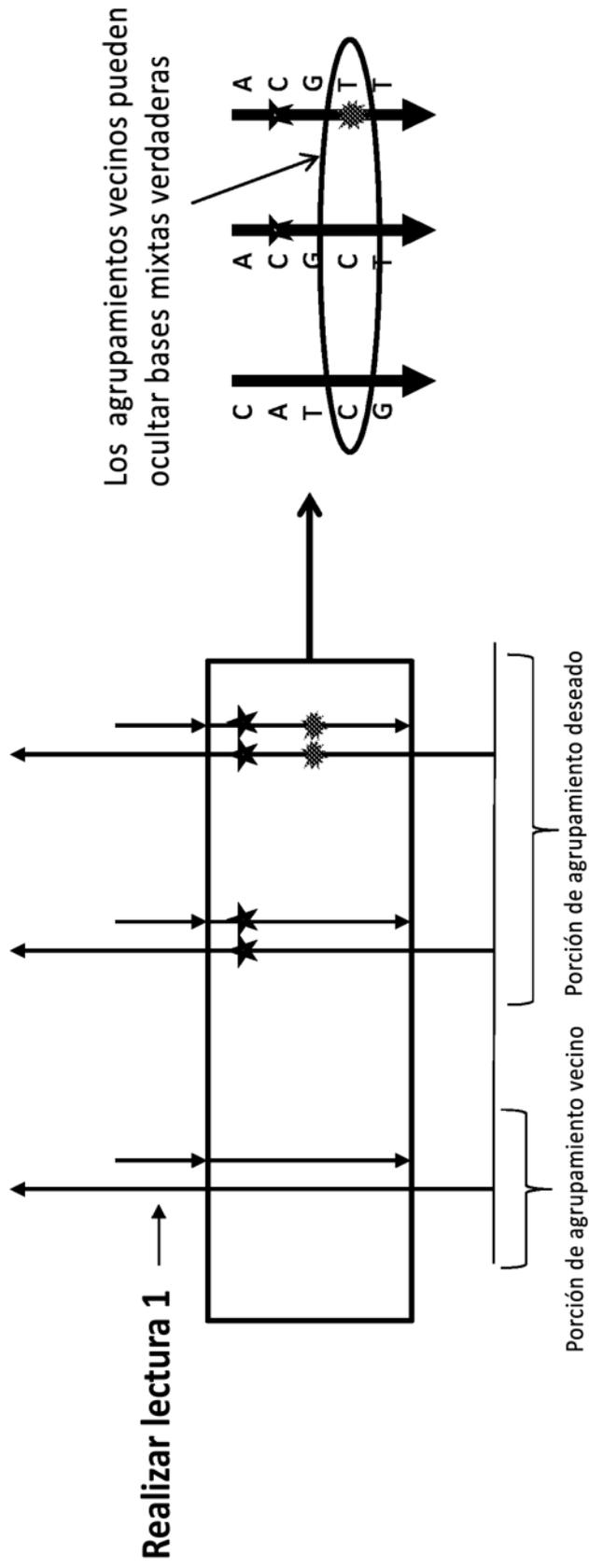


Figura 15



Estado de la técnica

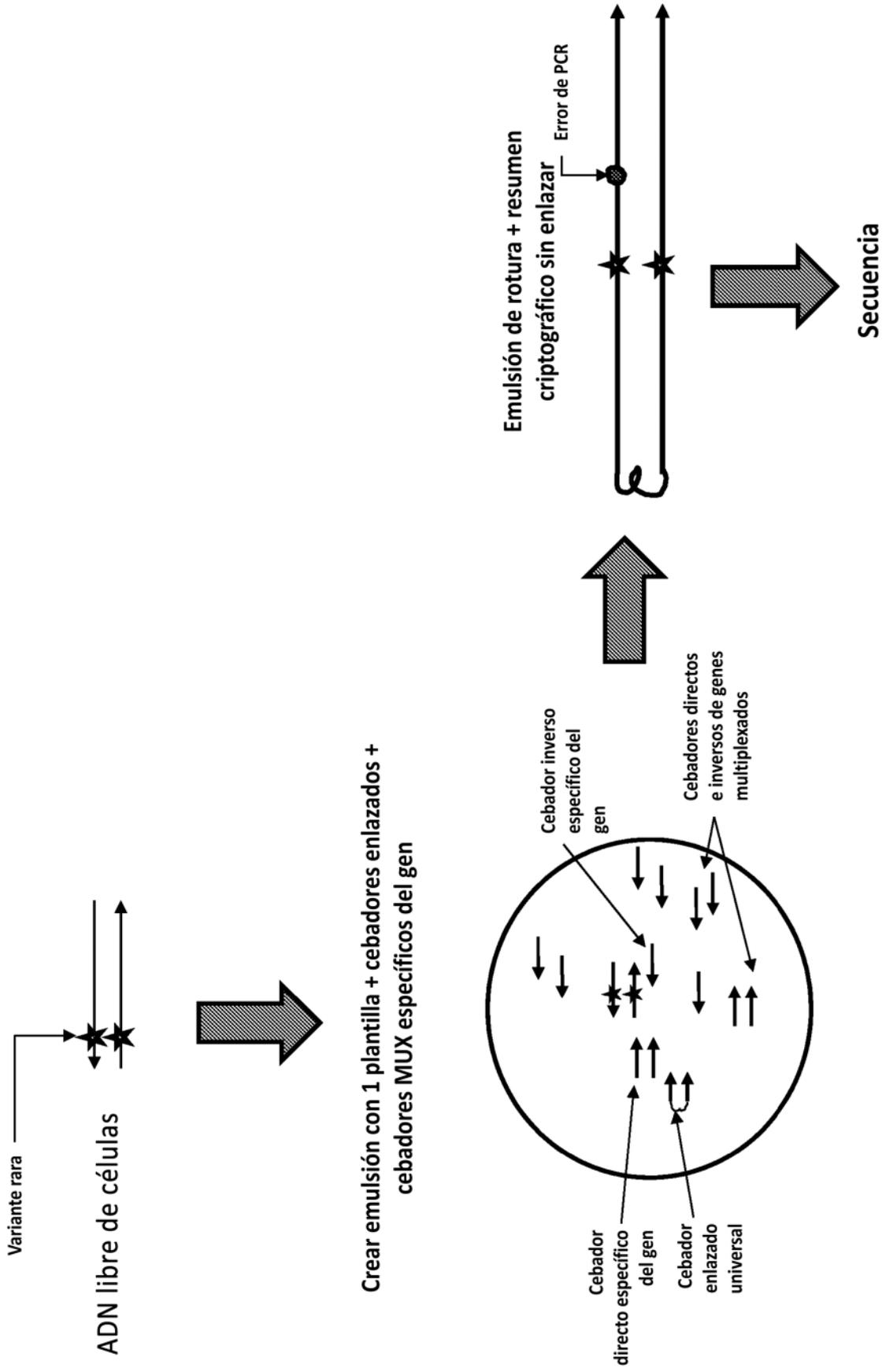


Figura 16

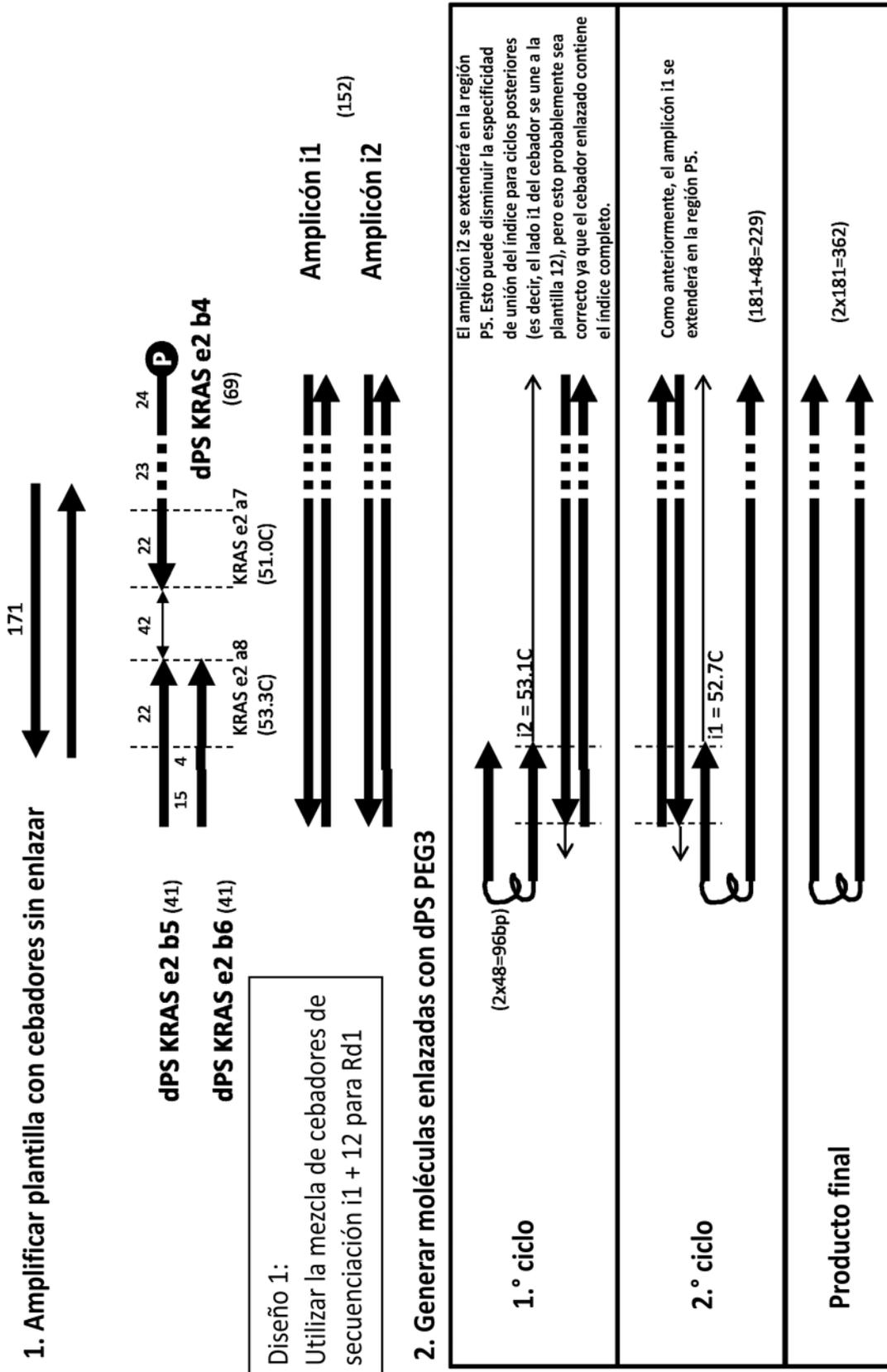
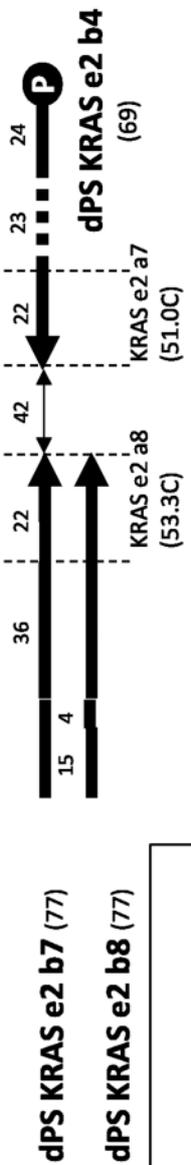


Figura 17

1. Amplificar plantilla con cebadores sin enlazar



Diseño 2:
Uso de una mezcla de cebador de
secuenciación independiente para Rd1



2. Generar moléculas enlazadas con dPS PEG3

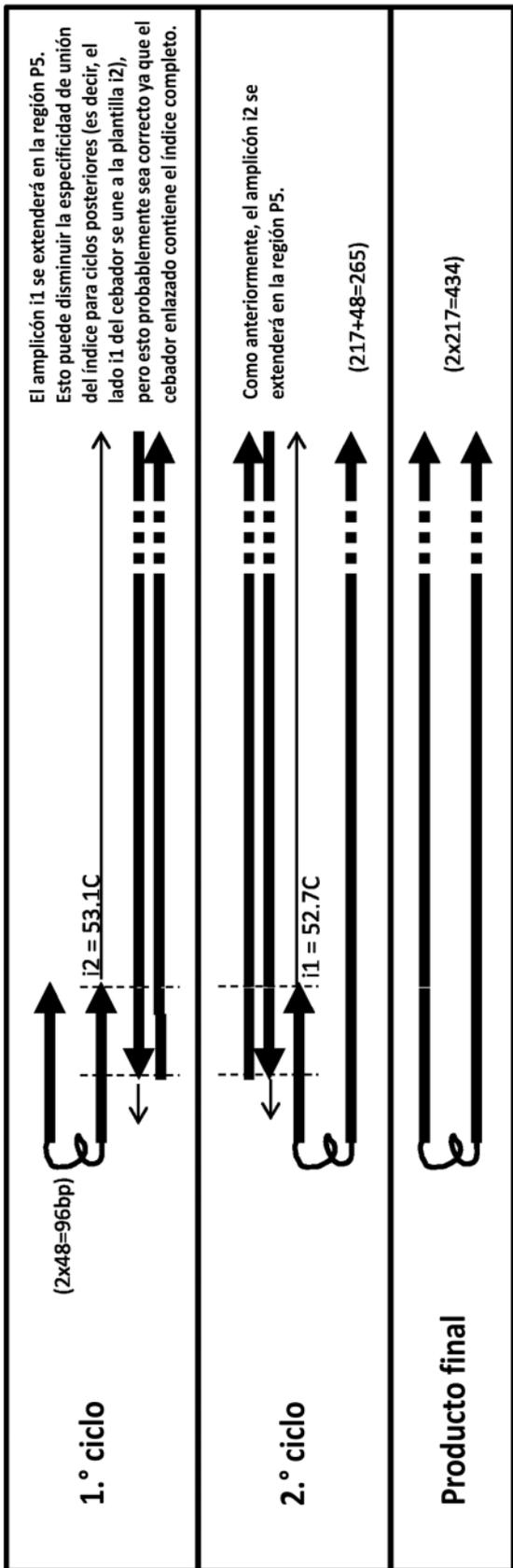


Figura 18

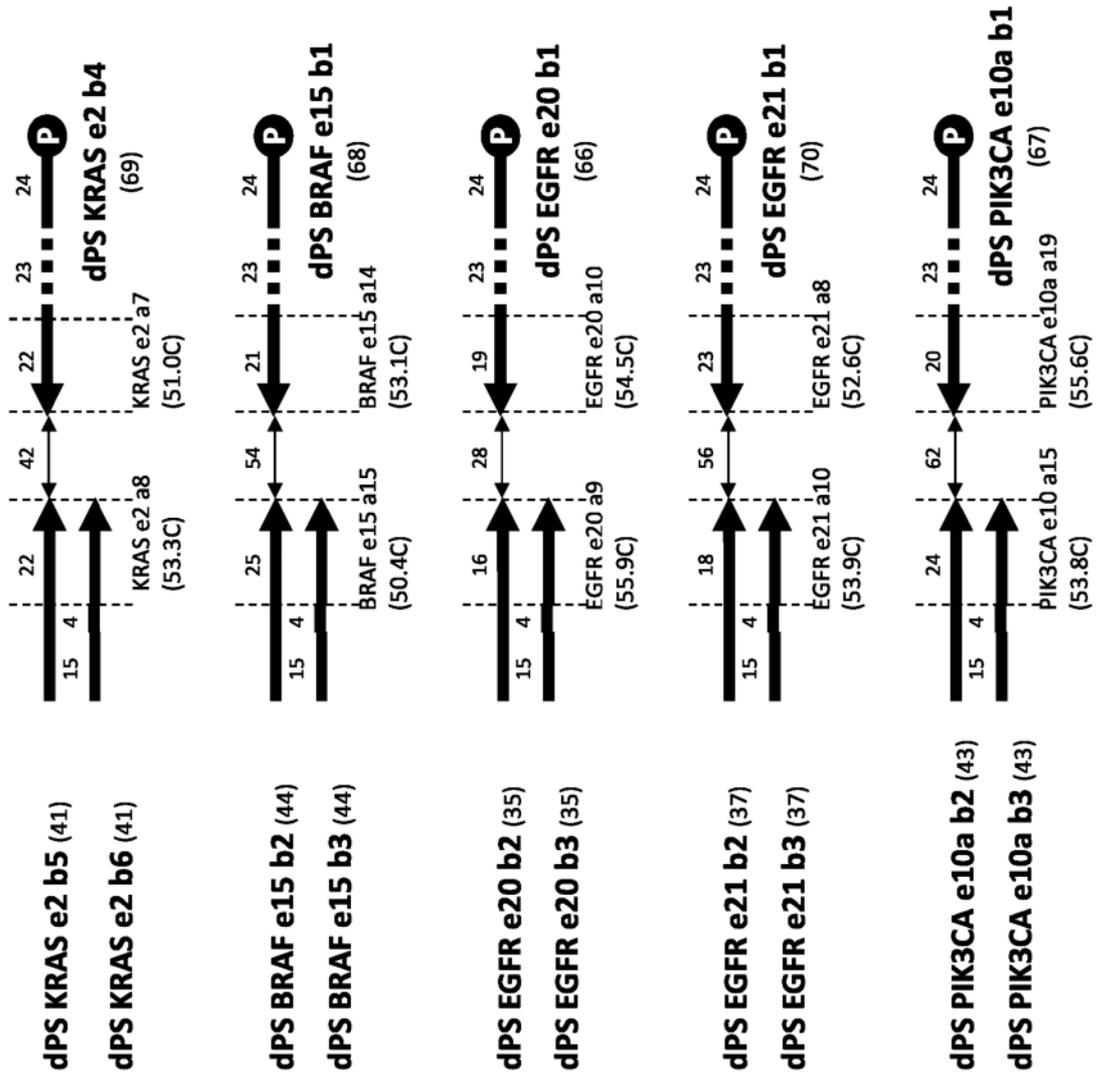


Figura 19

Región de control del índice: Espaciador de base 4
 I1: 5'-CTTCTTCTGCTGCAA CTGA-3'
 I2: 5'-TCCCTCCTATCATGG ACAC-3'

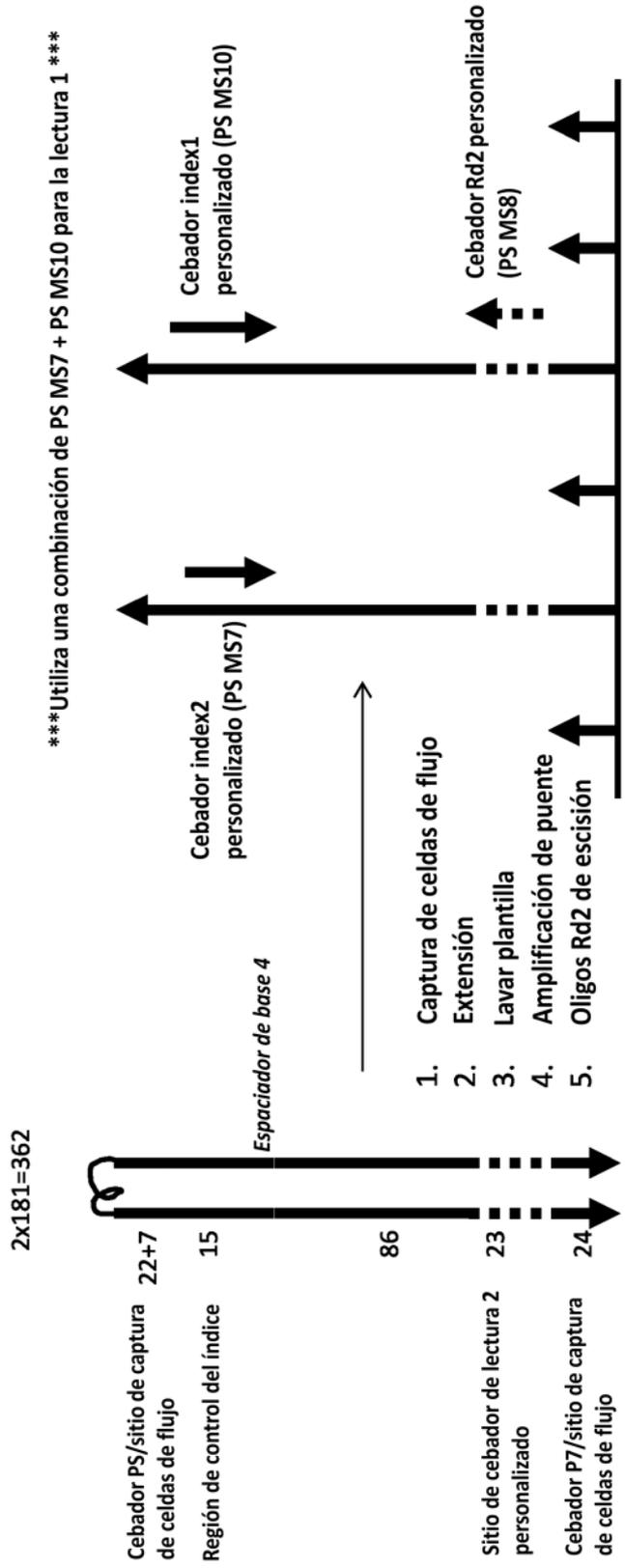


Figura 20

Región de control del índice: CTGA-3'
I1: 5'-CTTCTTCTGCTGCAA CTGA-3'
I2: 5'-TCCCTCCTATCATGG ACAC-3'

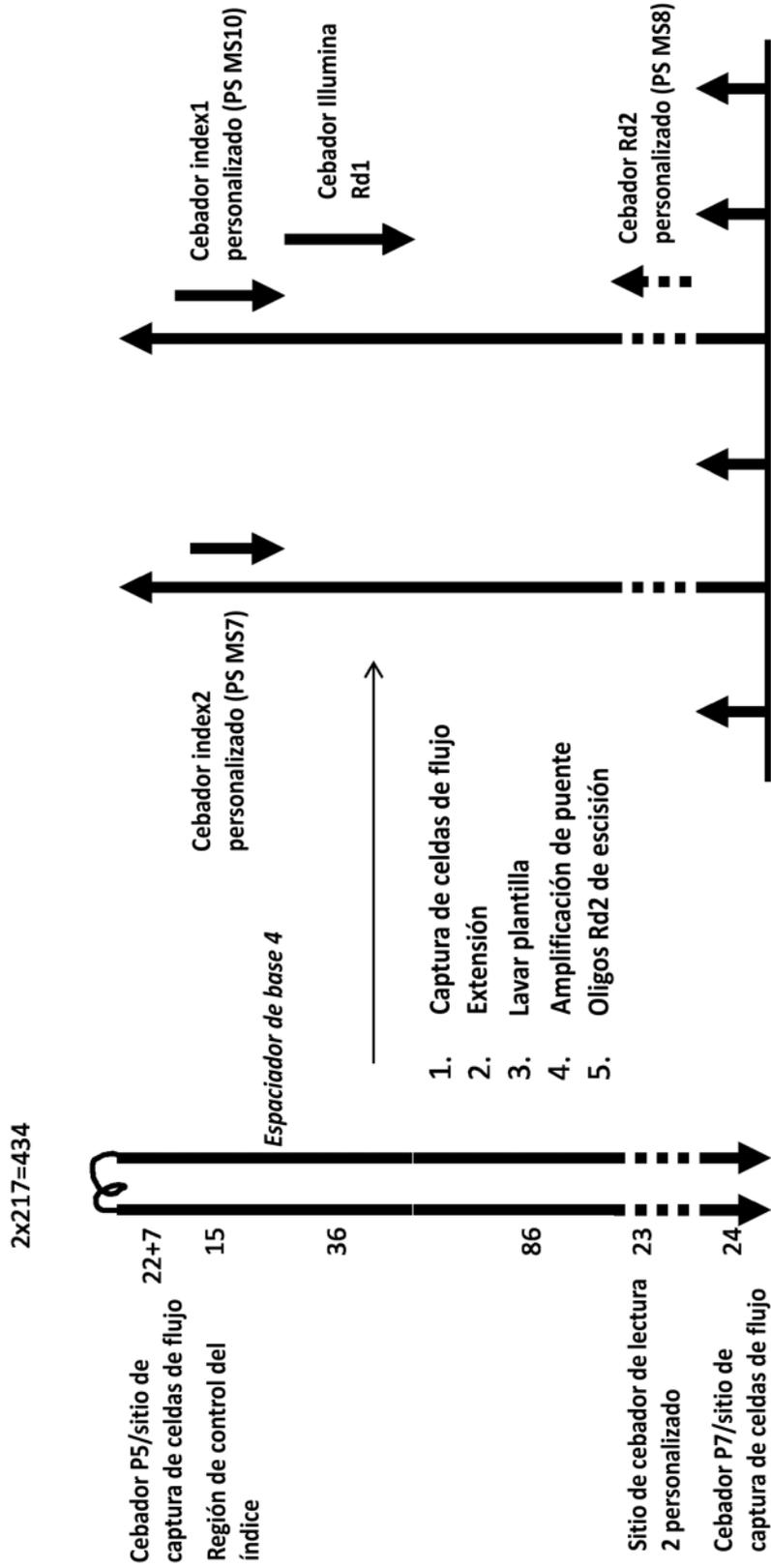


Figura 21

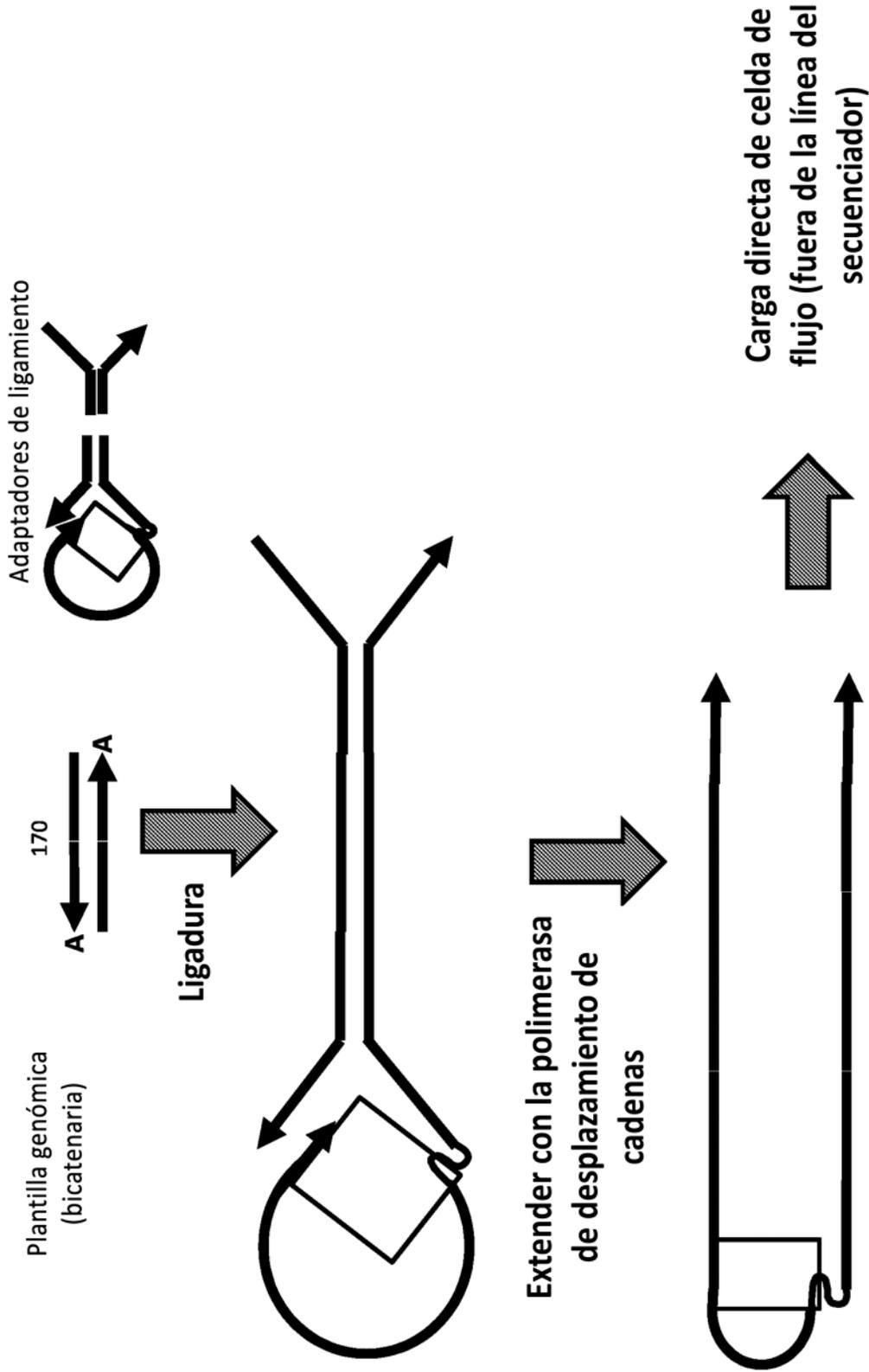
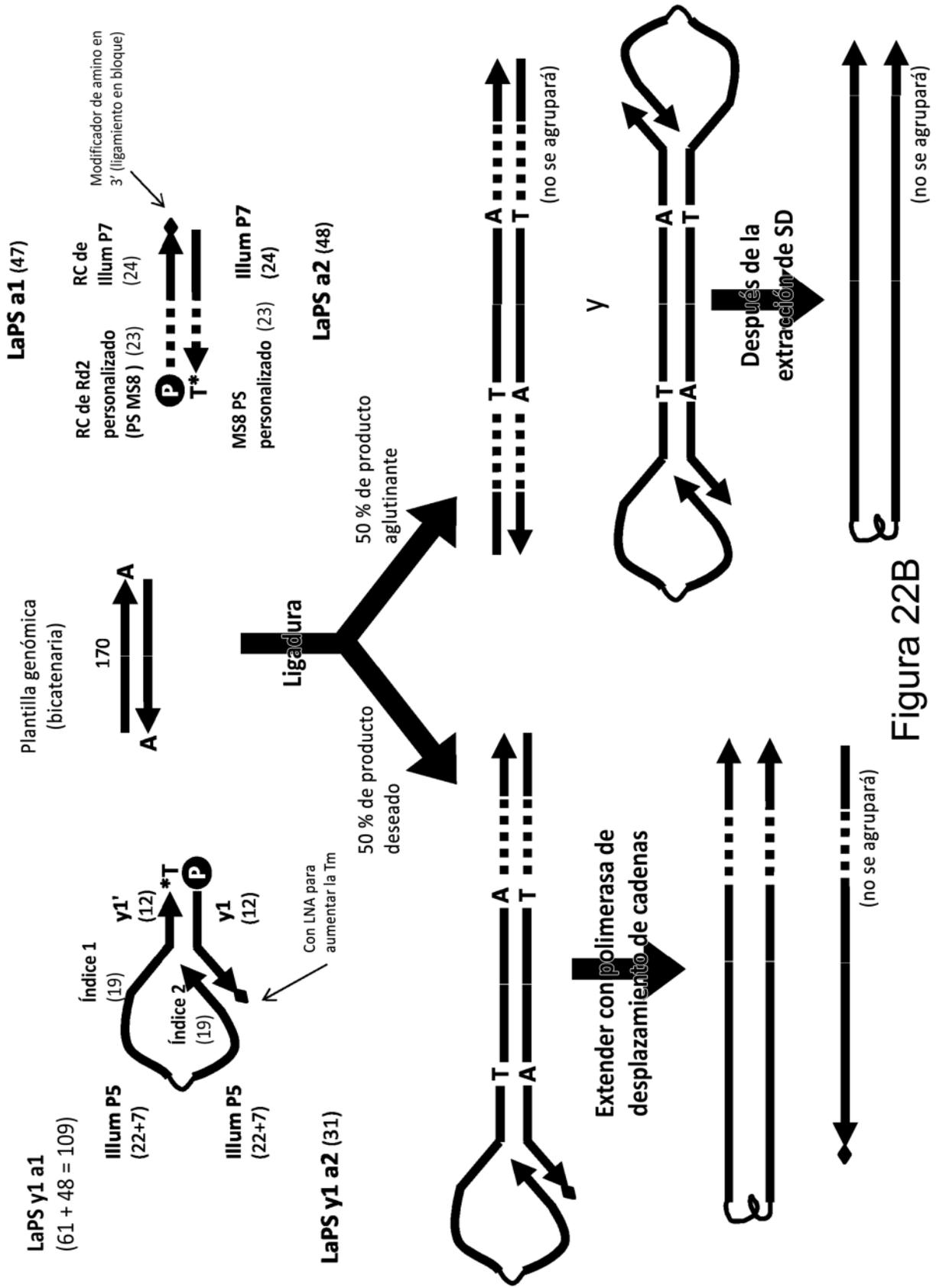


Figura 22A



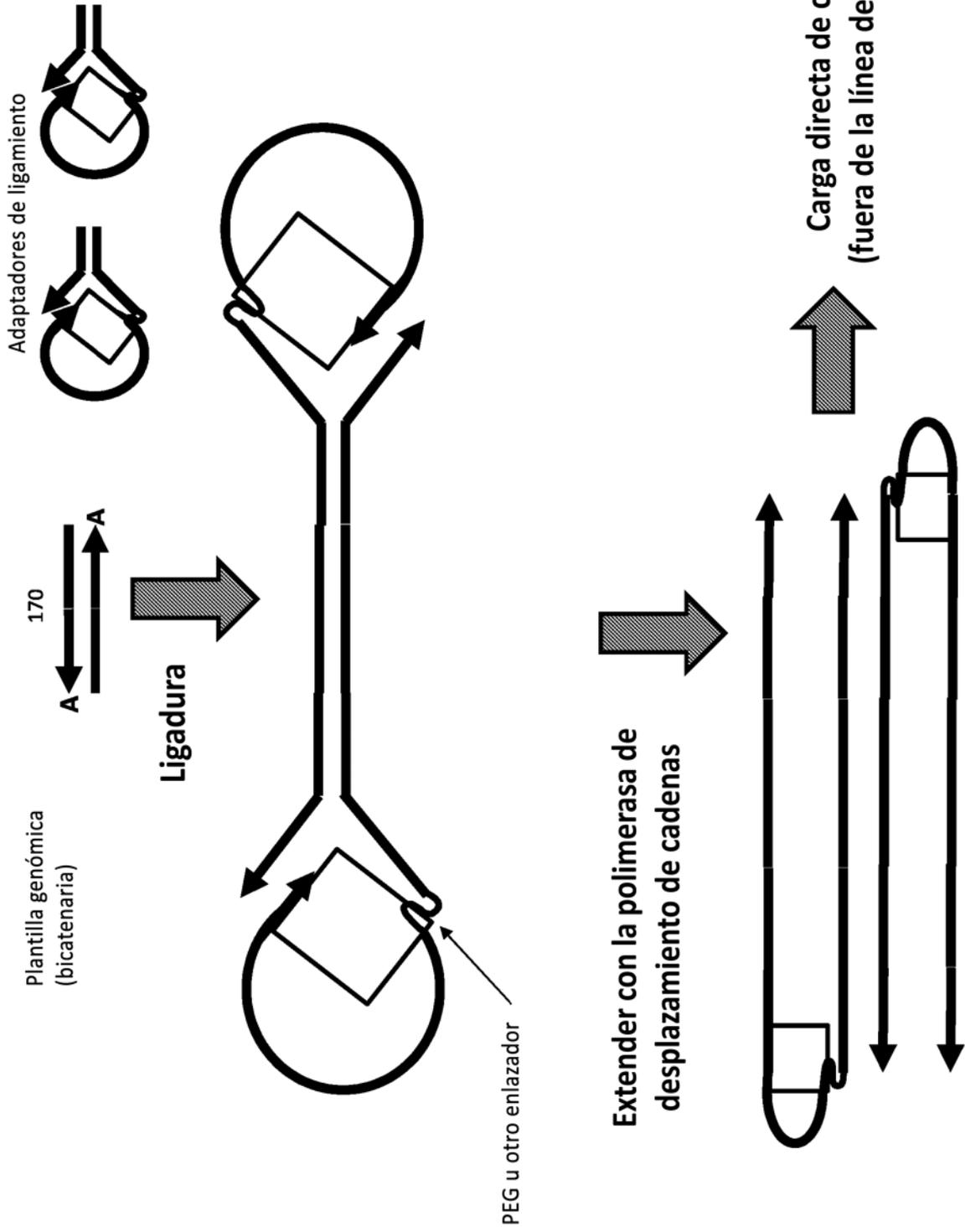


Figura 23A

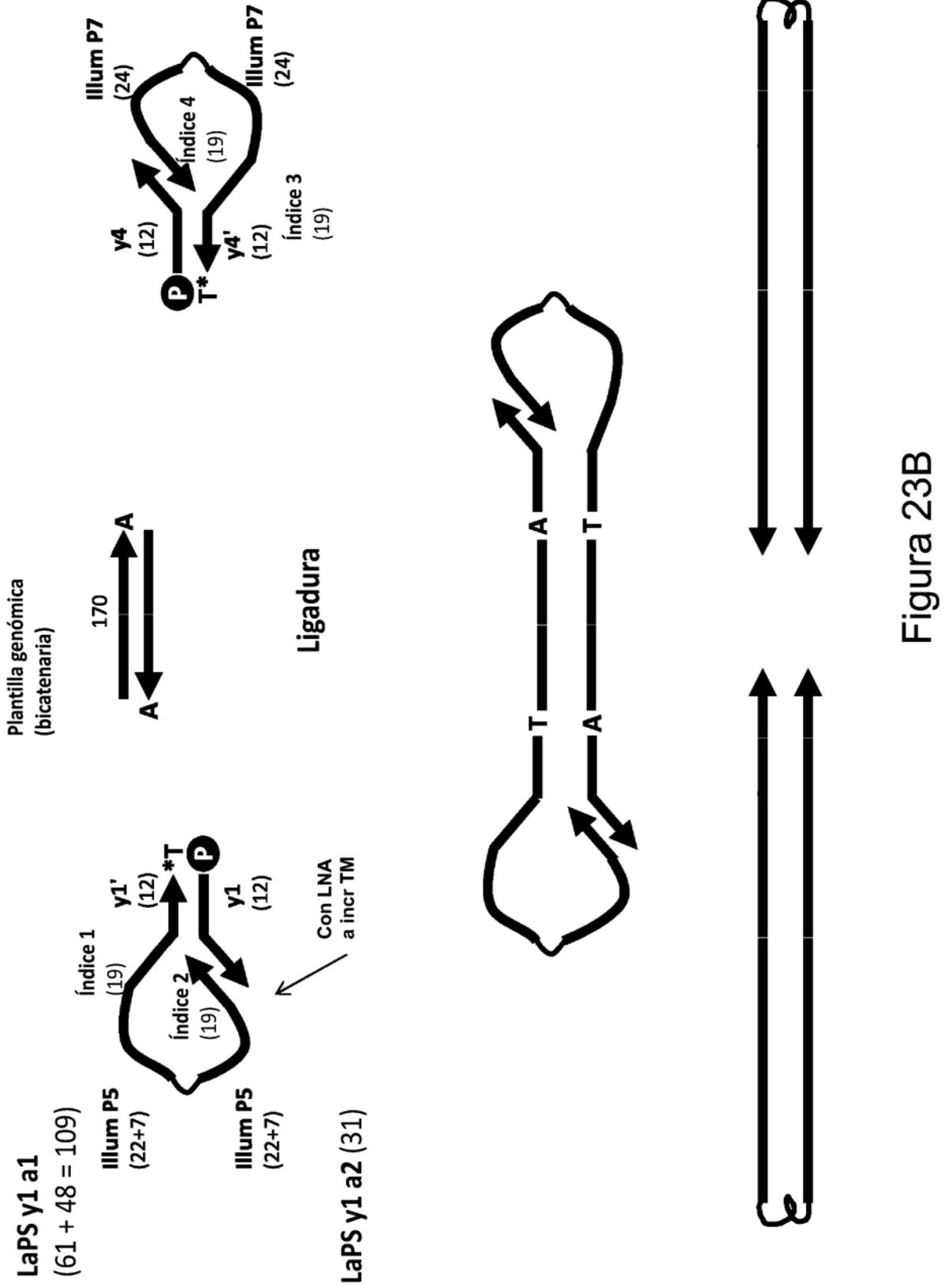


Figura 23B

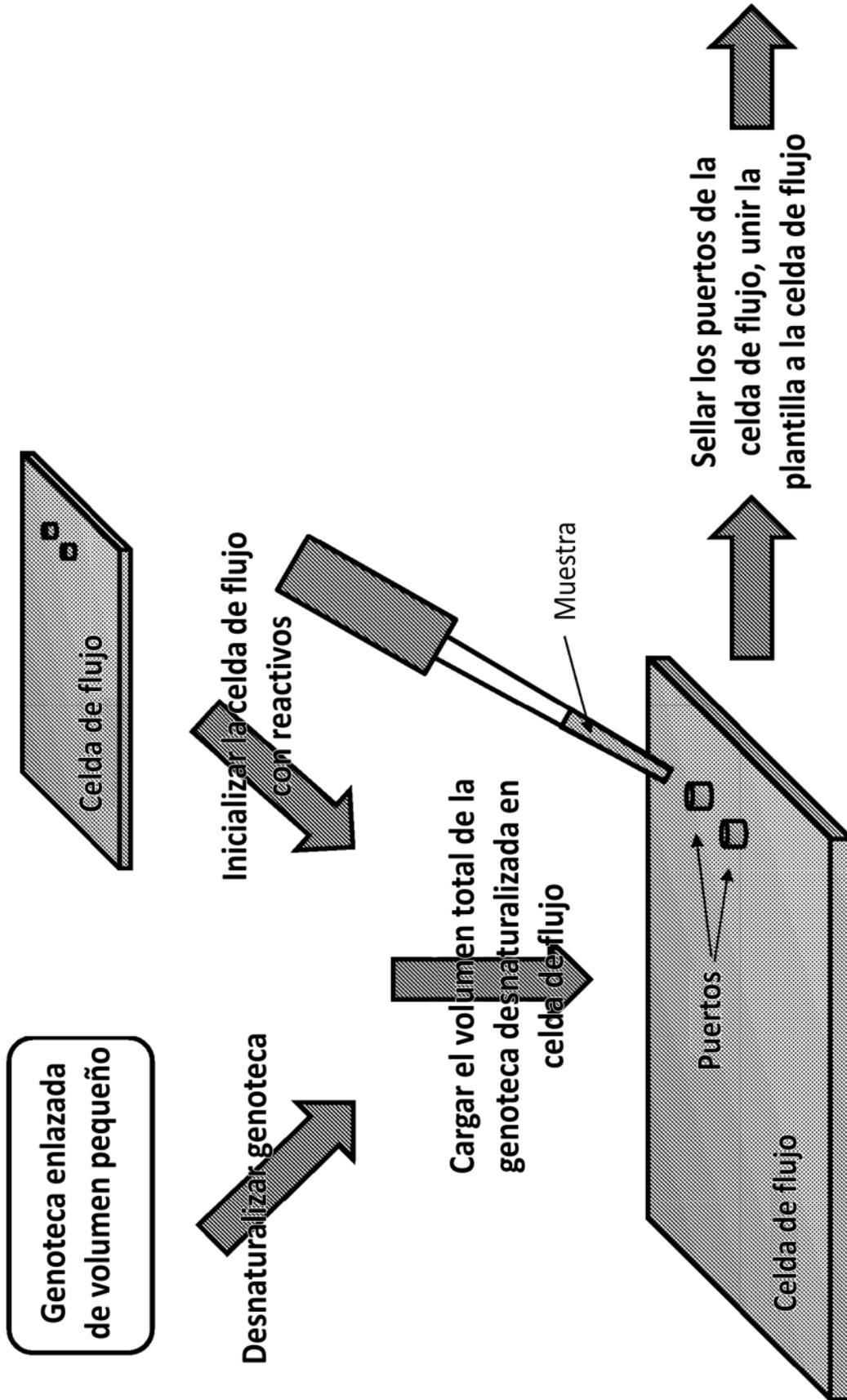


Figura 24A

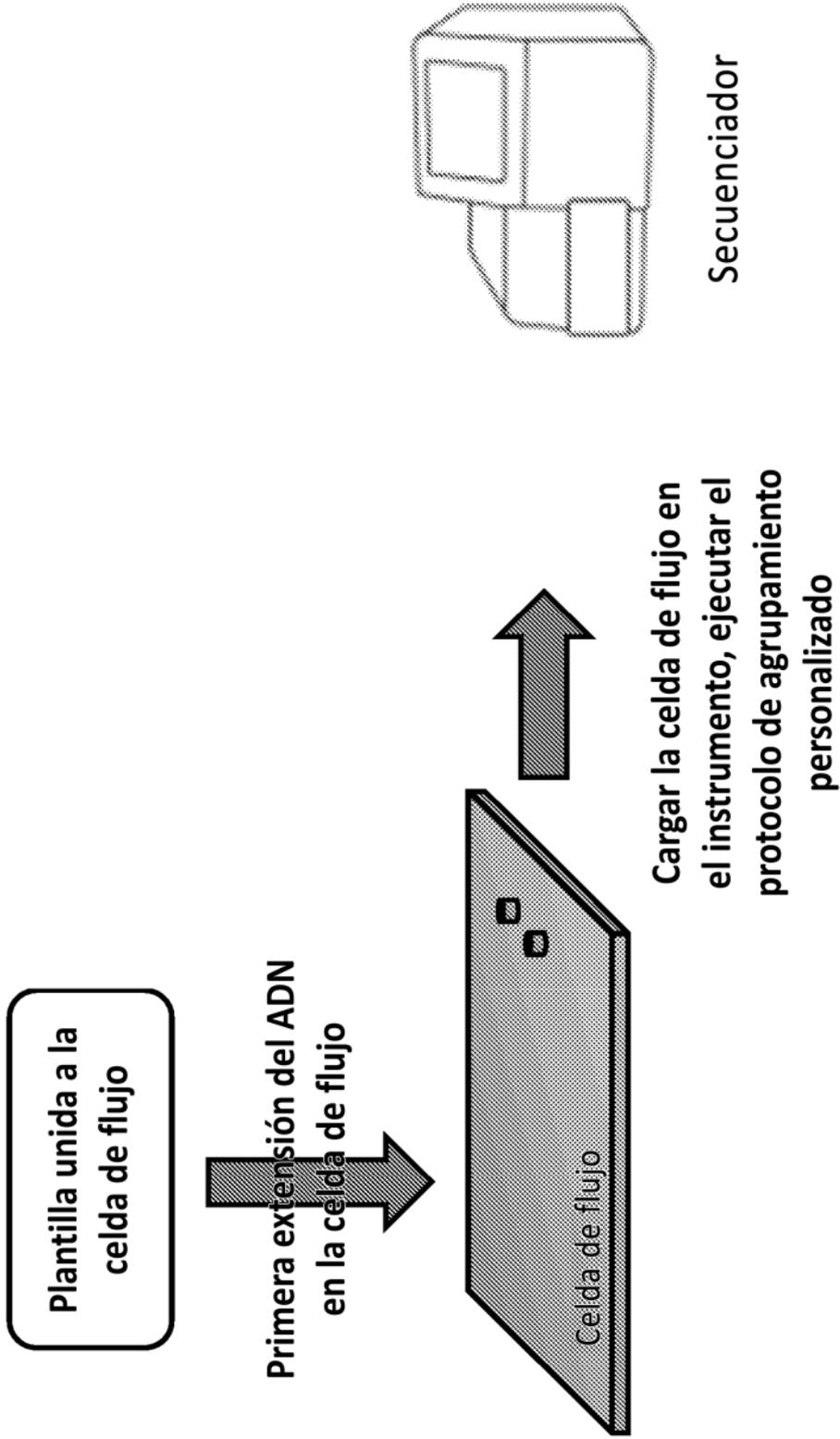


Figura 24B

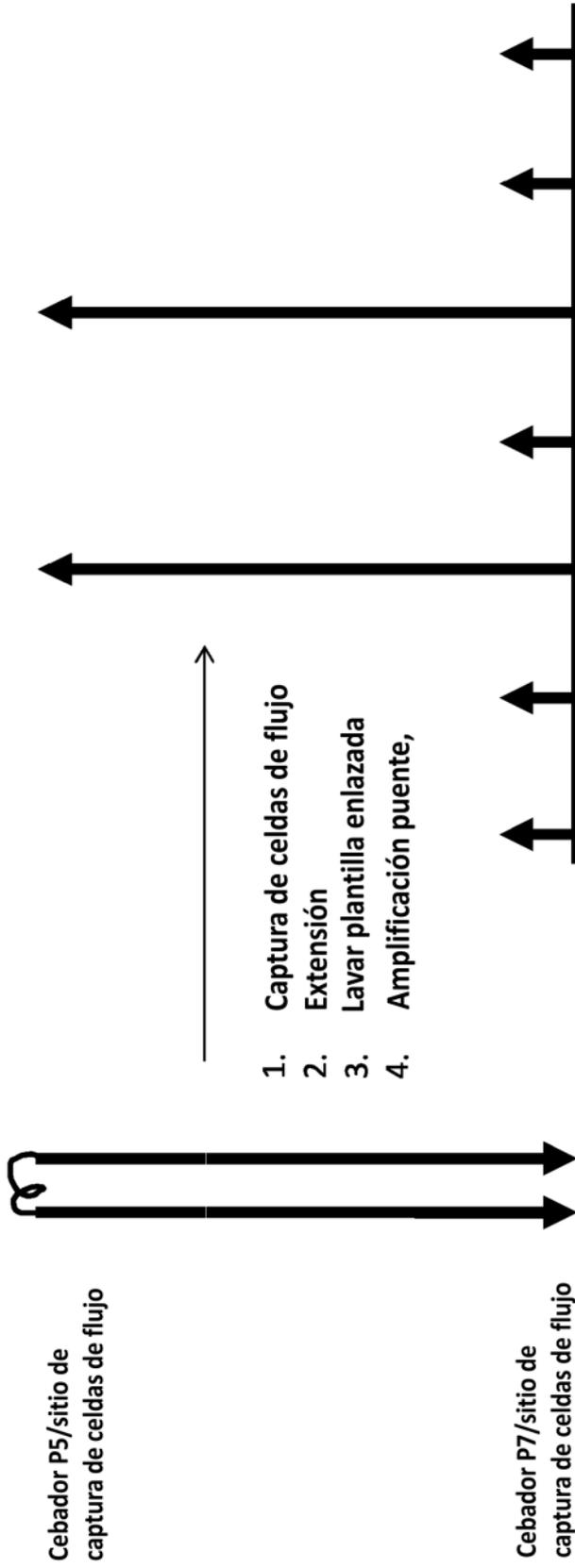


Figura 25

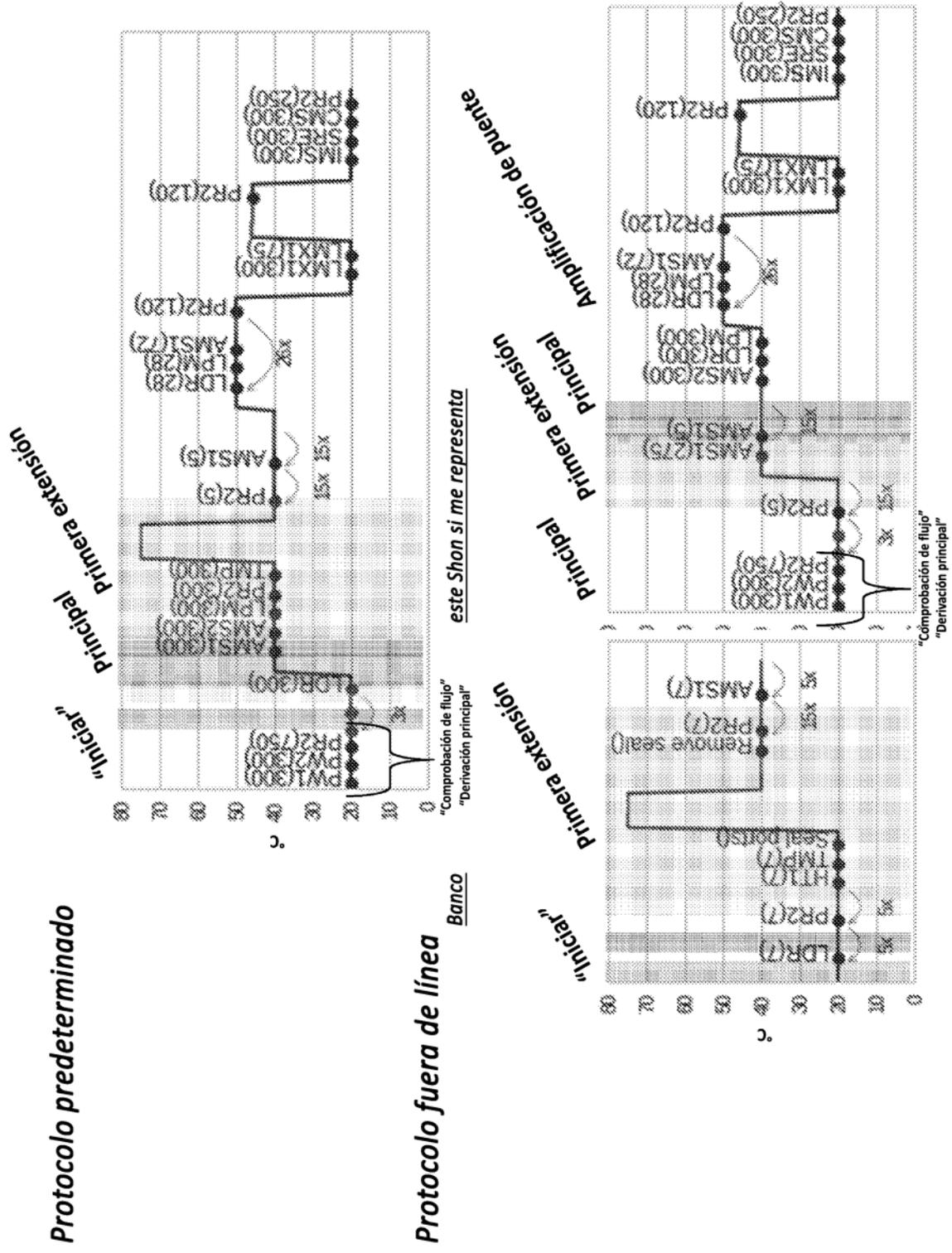


Figura 26

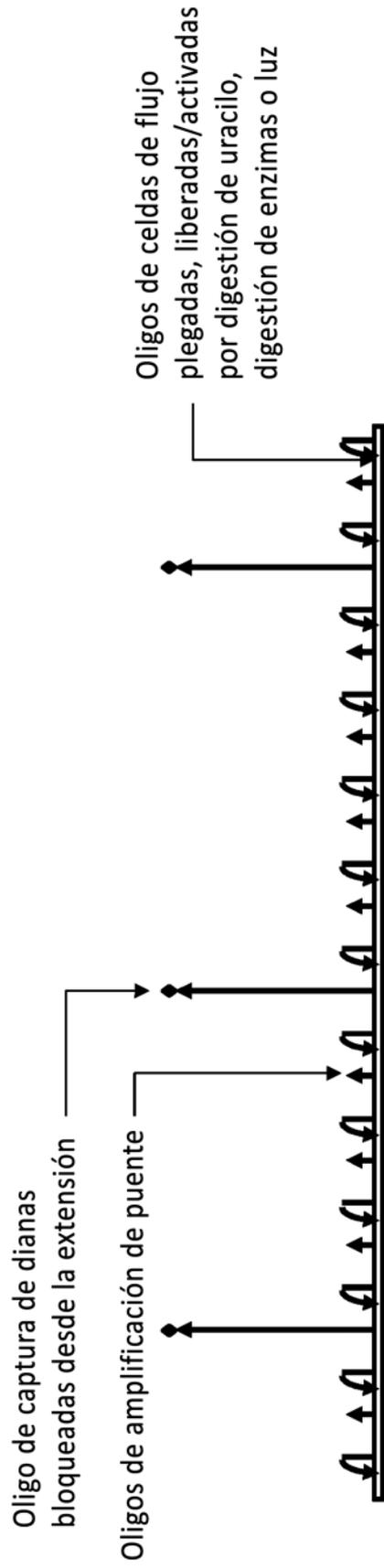


Figura 27

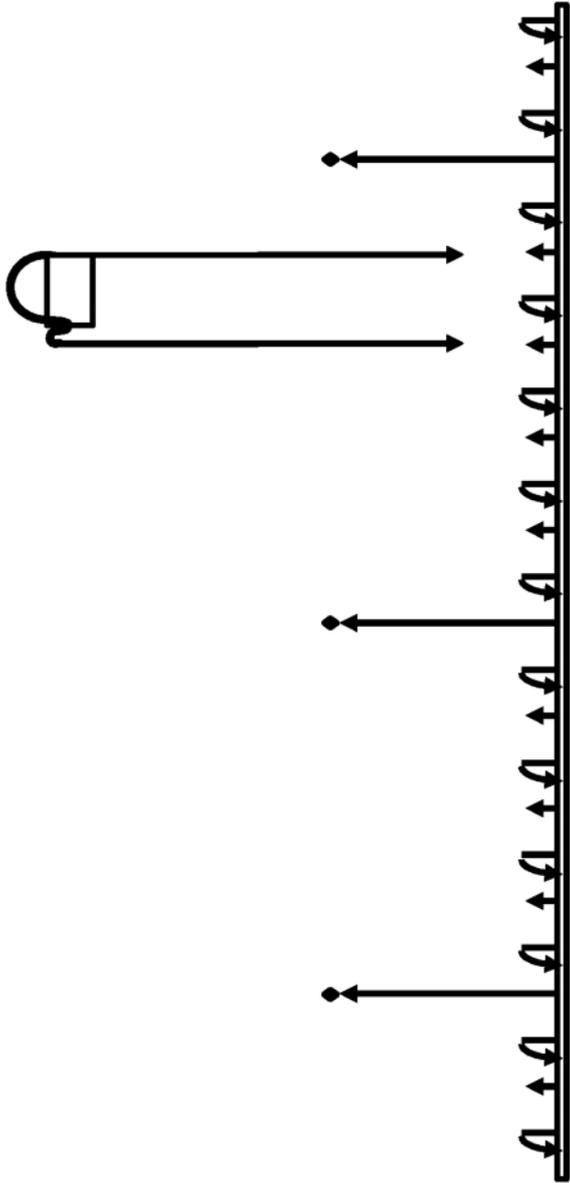


Figura 28A

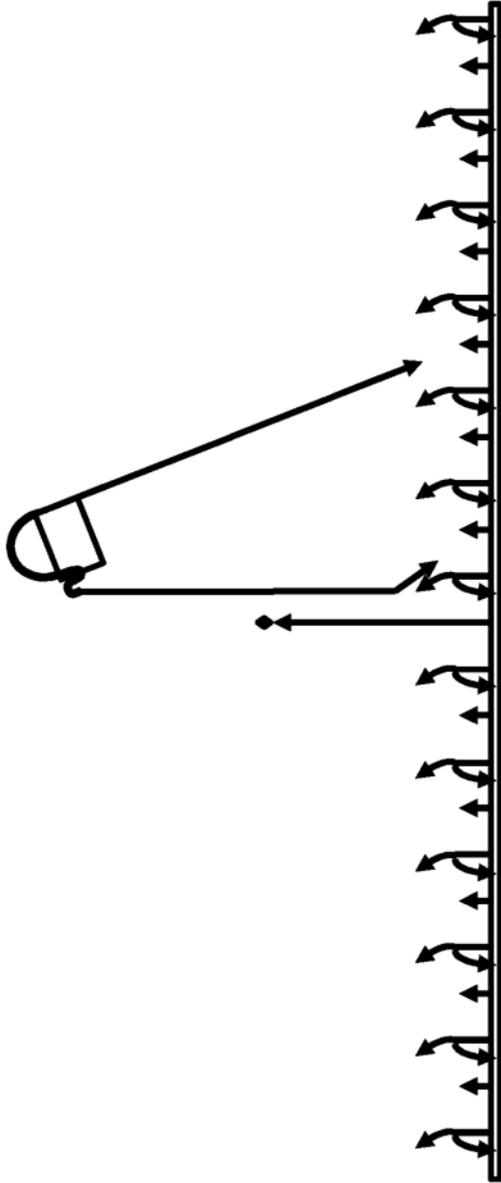


Figura 28B

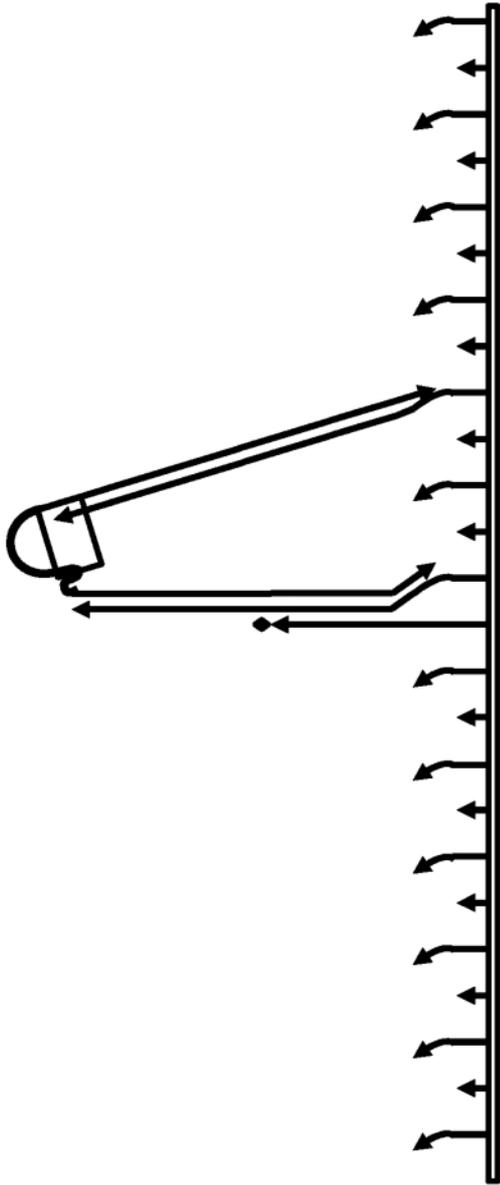


Figura 28C

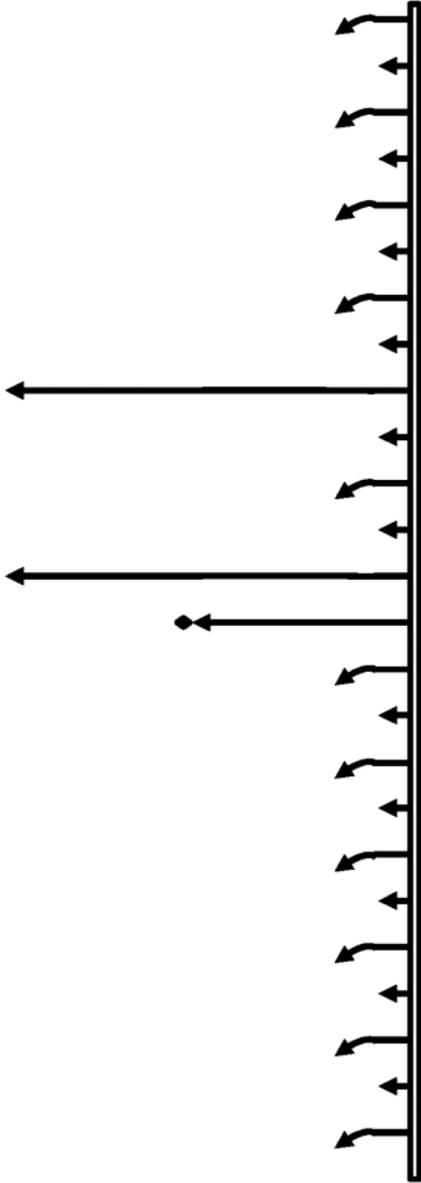


Figura 28D

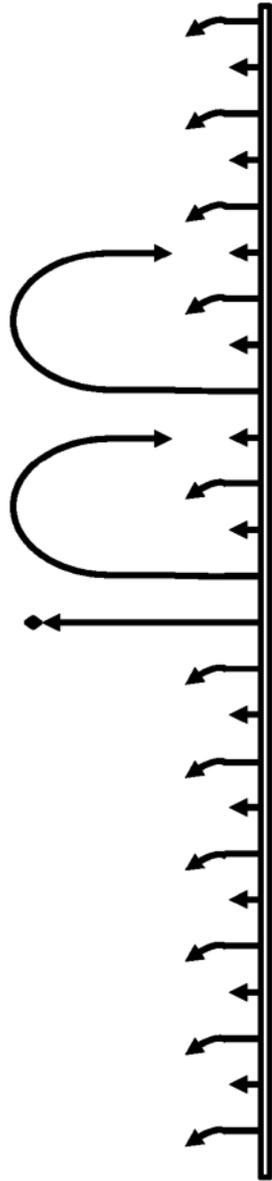
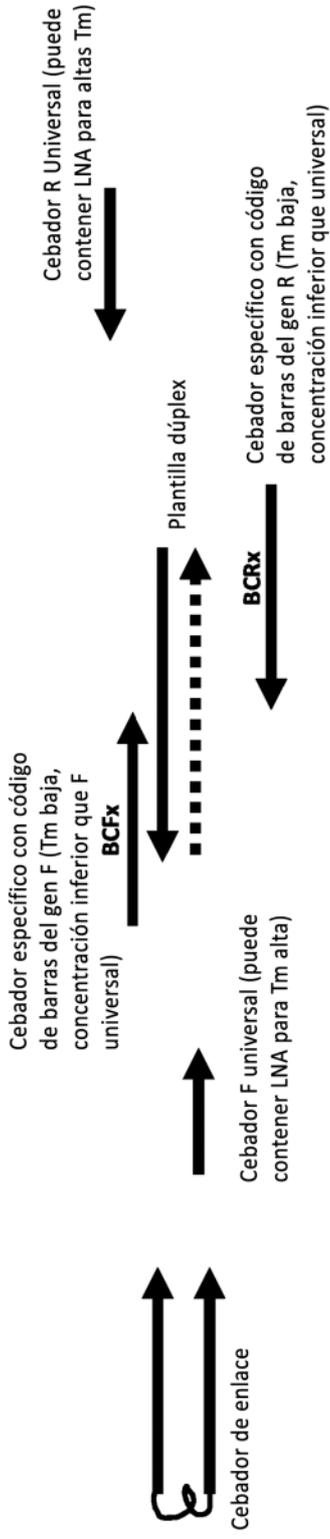


Figura 28E



Ciclo 1

Temperatura de hibridación baja (permite cebadores codificados para unirse a la plantilla)



Ciclo 2

Temperatura de hibridación baja

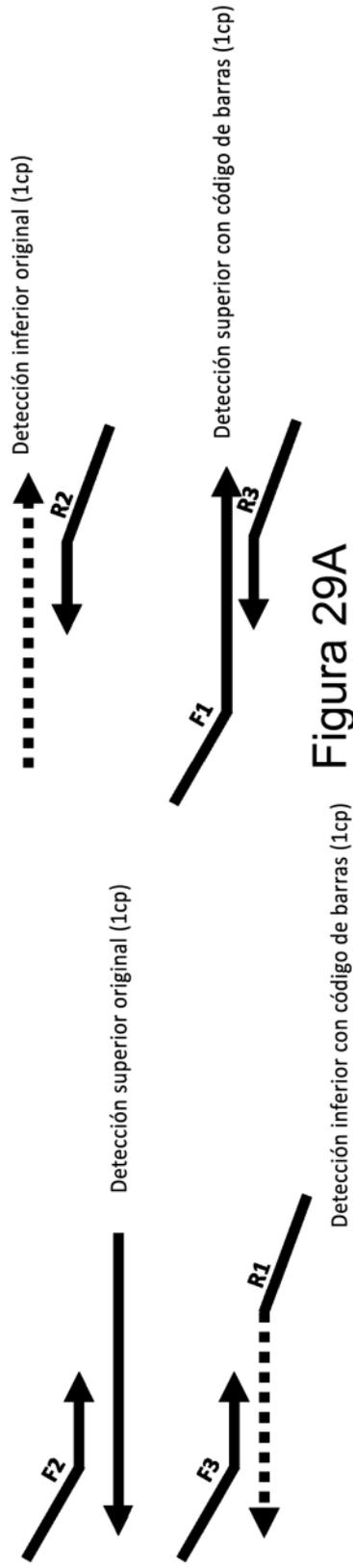
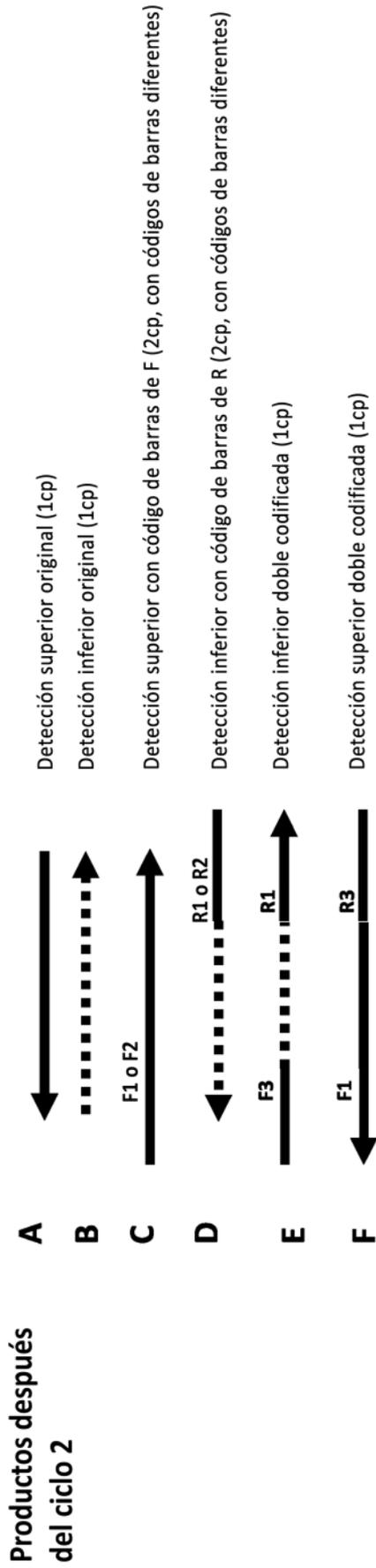


Figura 29A

Detección inferior con código de barras (1cp)

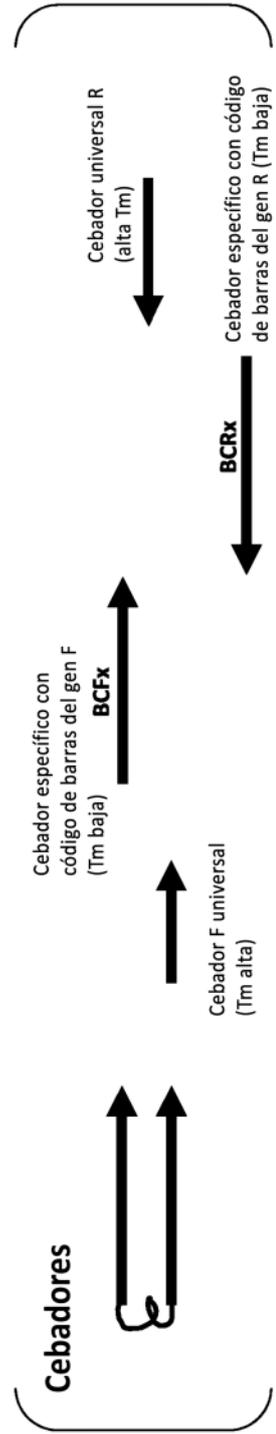


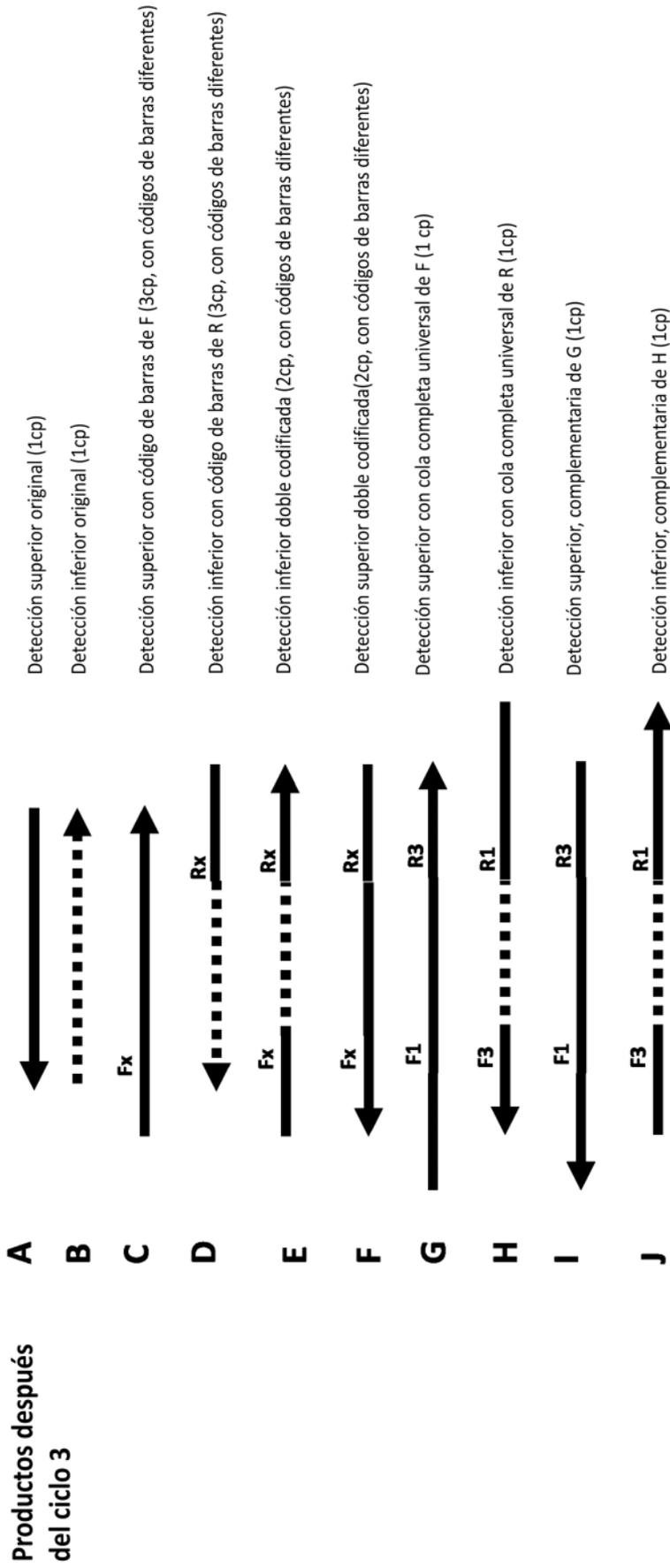
Ciclo 3

Temperatura de hibridación baja. Permitir el 1.º ciclo de unión del cebador universal.



Figura 29B





Ciclo 4

Temperatura de hibridación baja. Permitir el 2.º ciclo de unión del cebador universal.



Figura 29C

Ciclo 5 - X

Temperatura de hibridación alta. Solo permitir la unión de cebadores universales a amplicones de cola completa universal



Últimos ciclos

Temperatura de hibridación baja. Permitir el enlace de cadenas amplificadas a través de una parte de la cola universal de F



Figura 29D

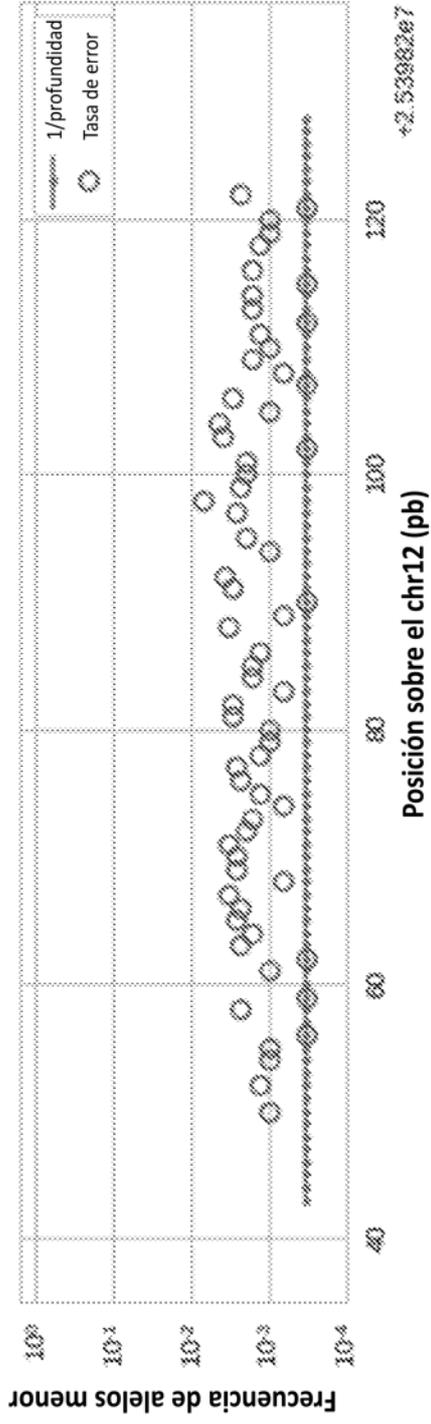


Figura 30

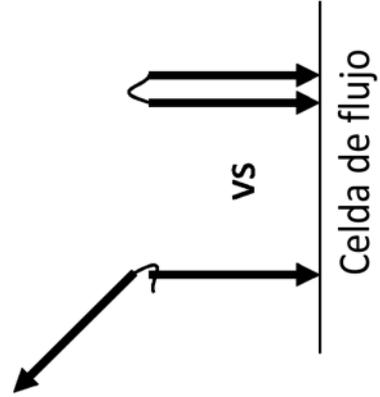


Figura 31

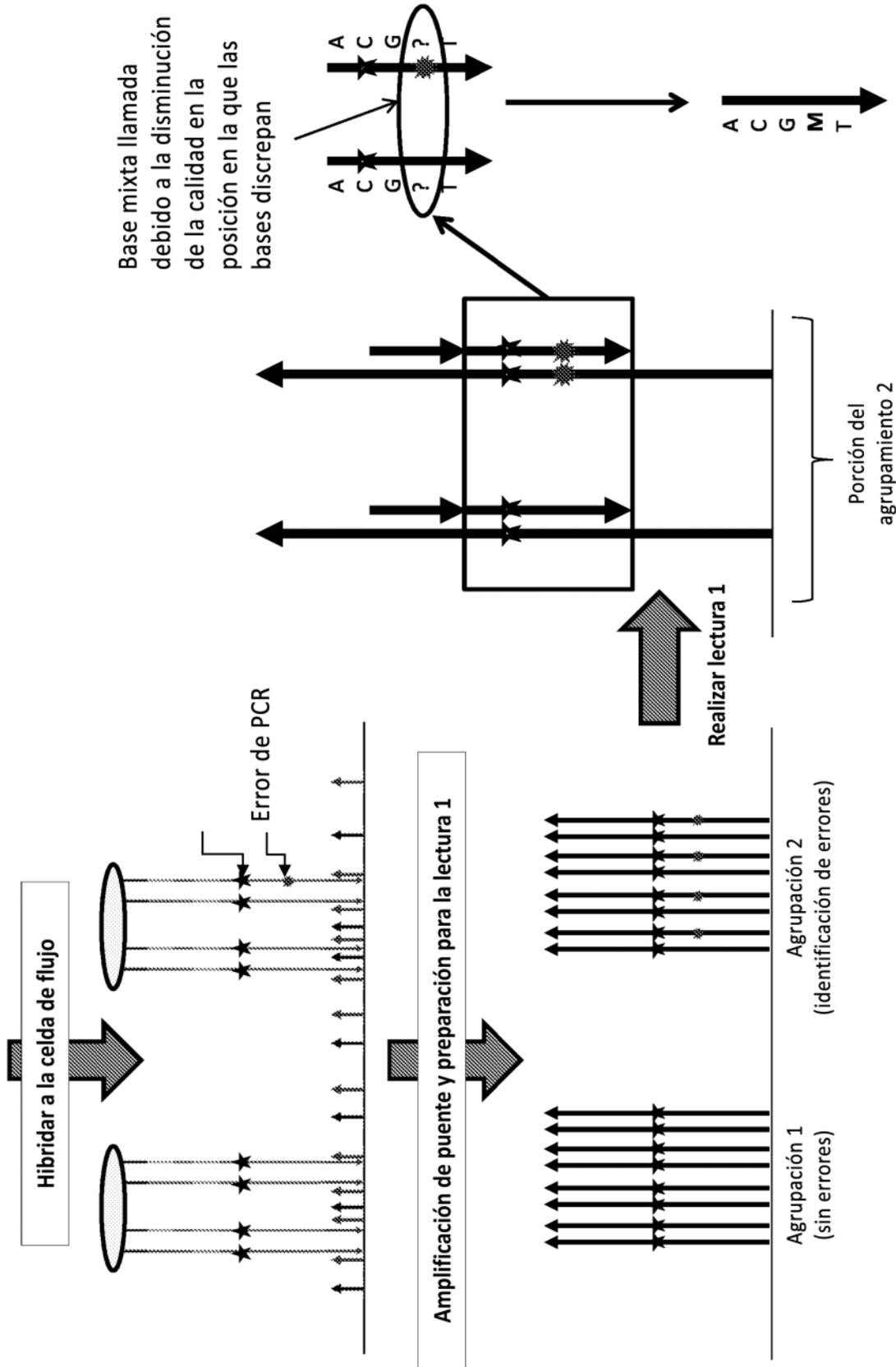


Figura 32

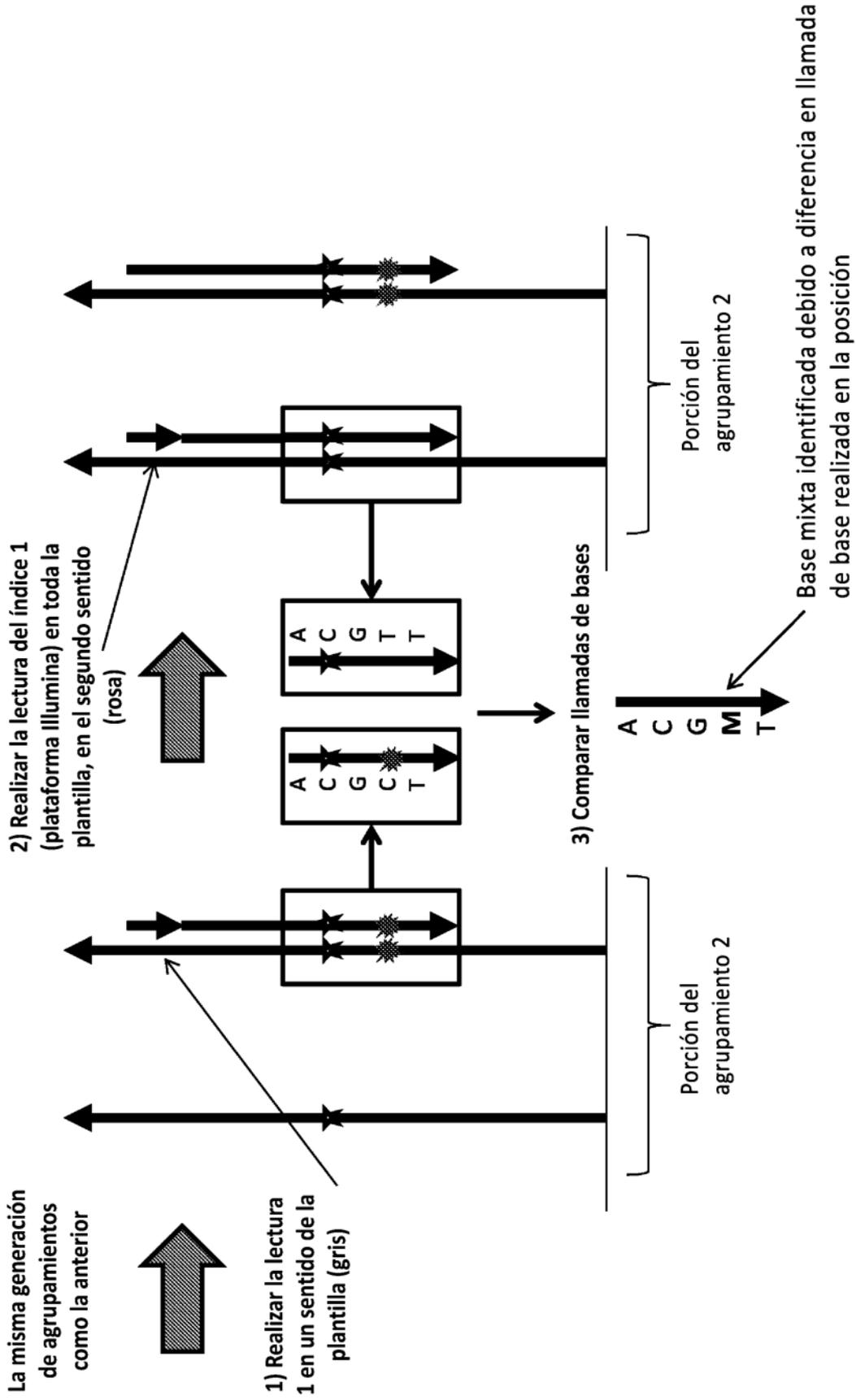


Figura 33

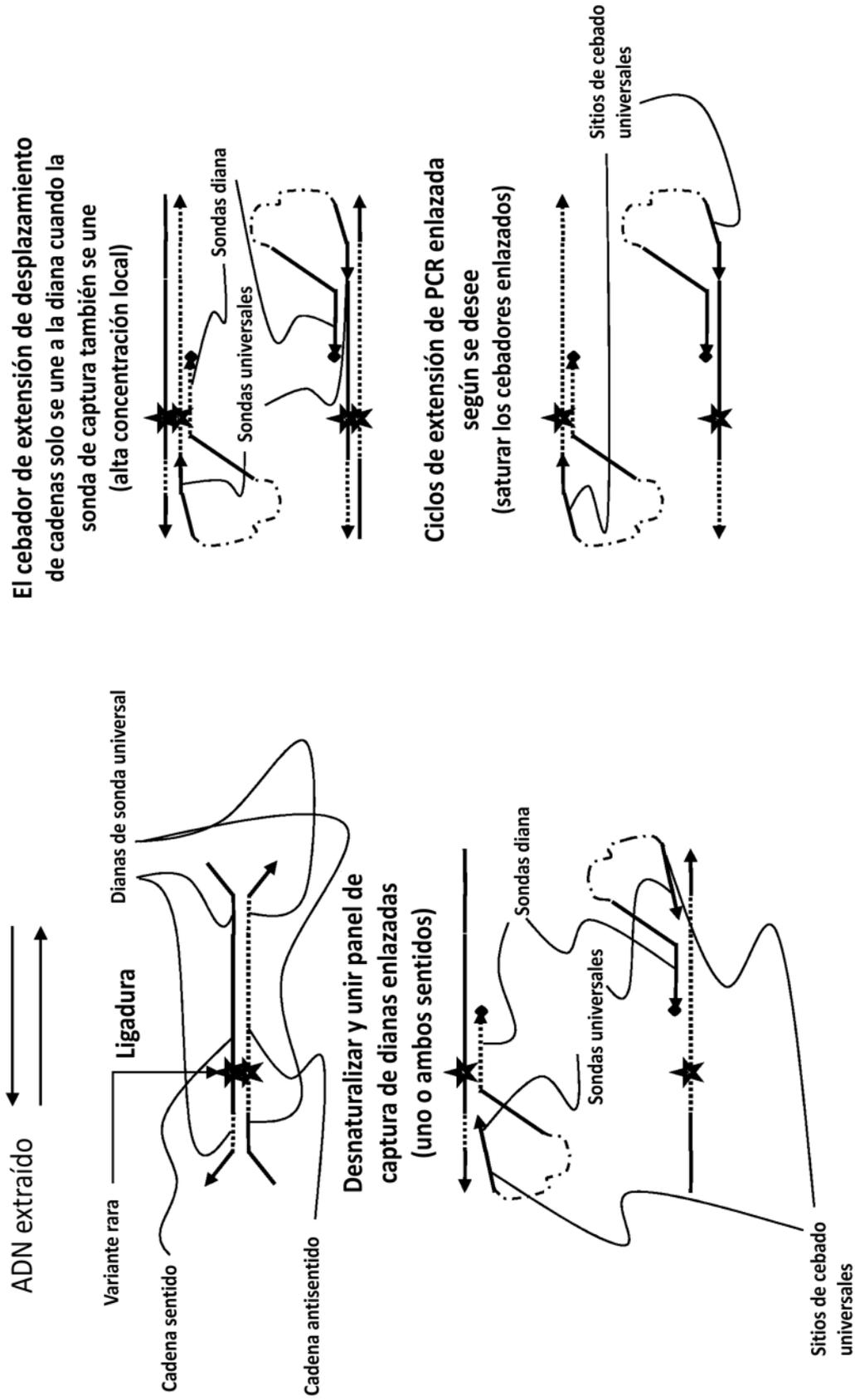


Figura 34

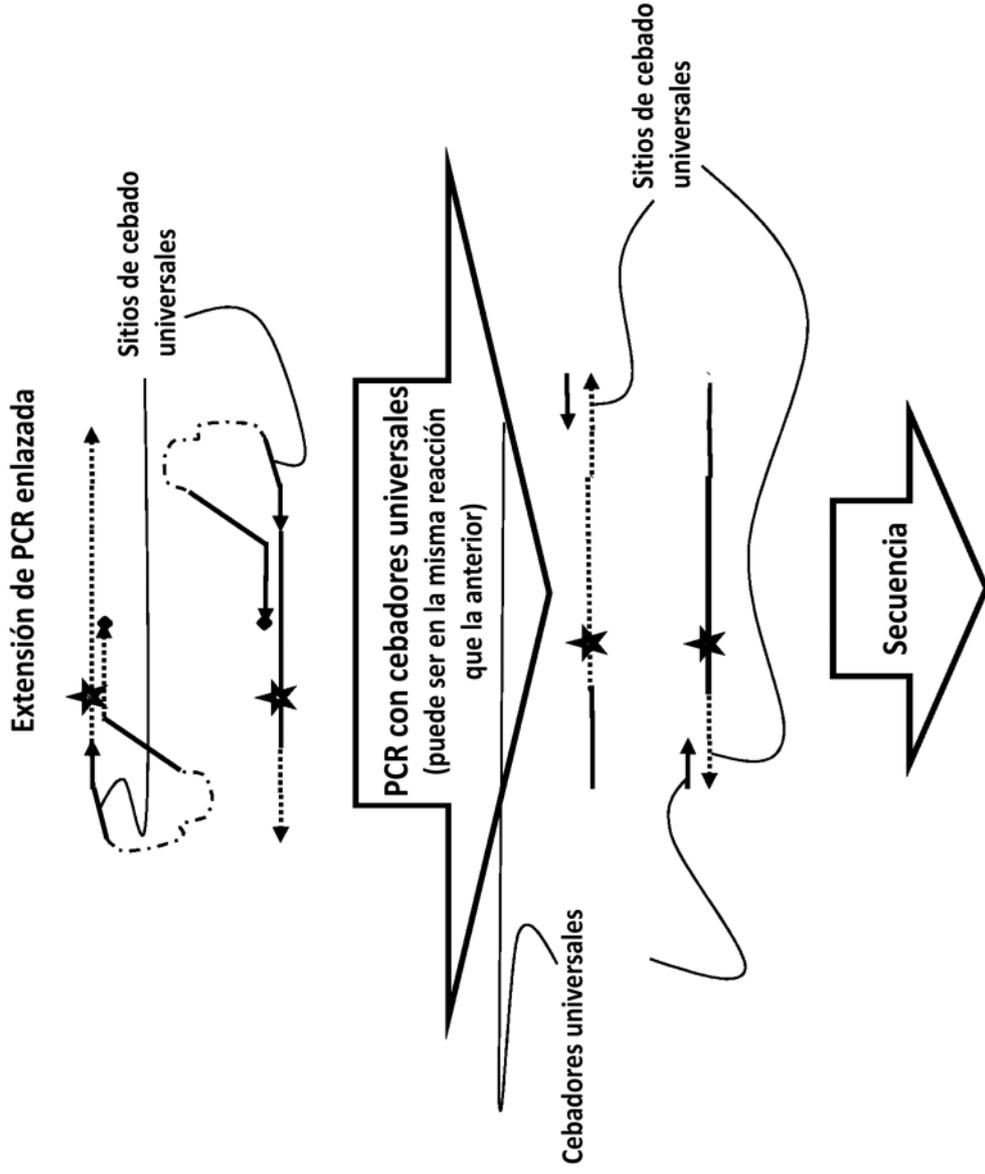


Figura 35

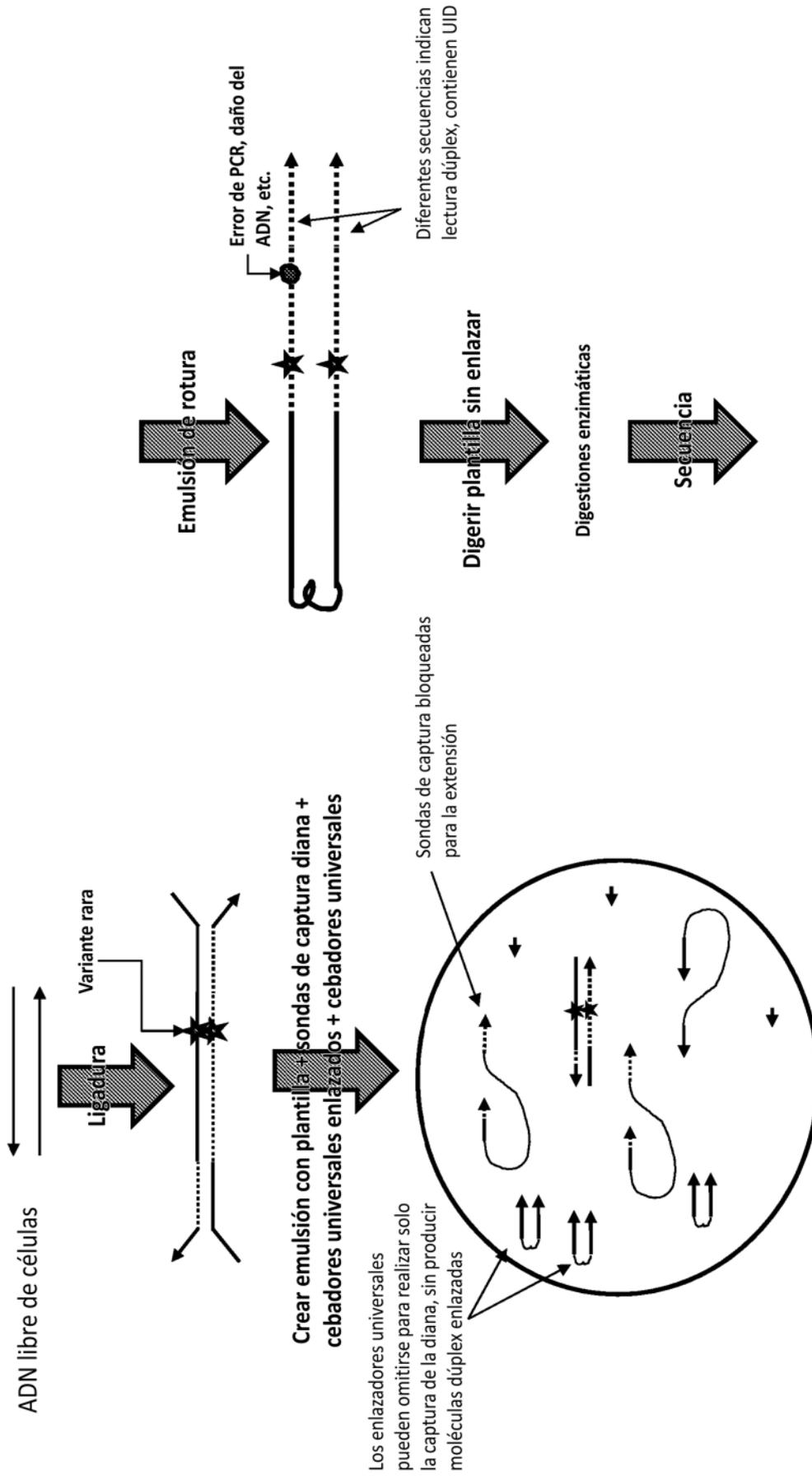


Figura 36

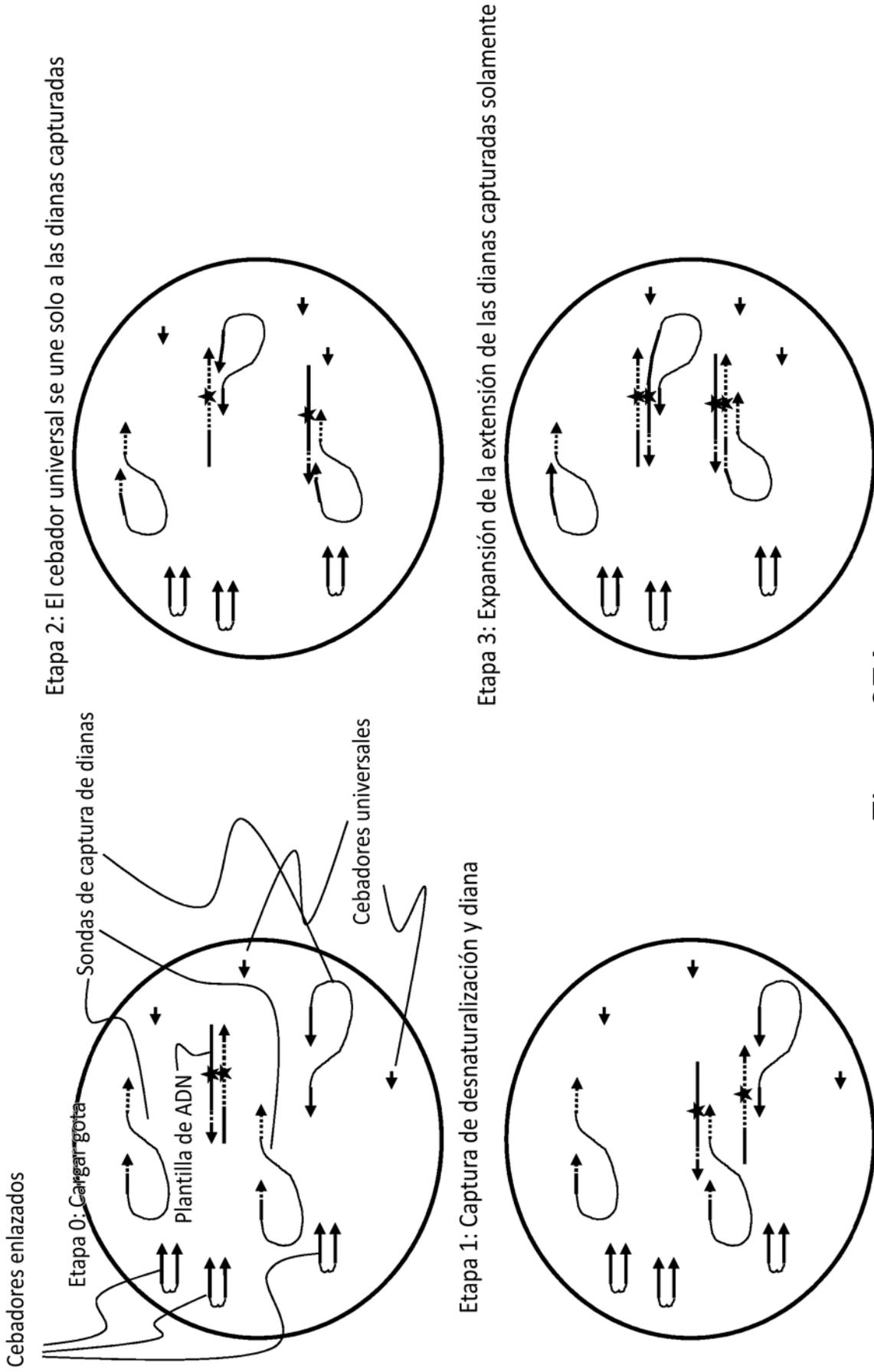
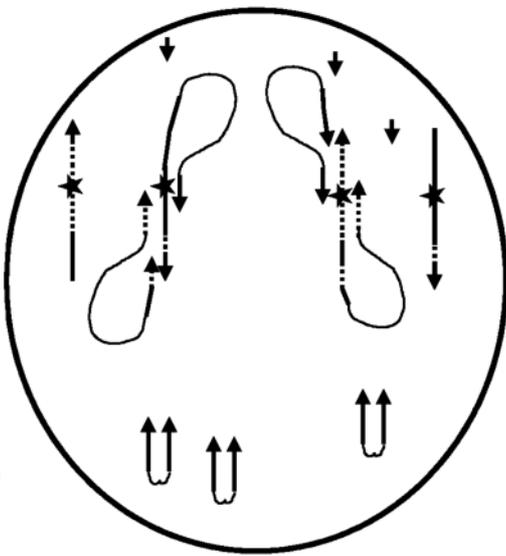
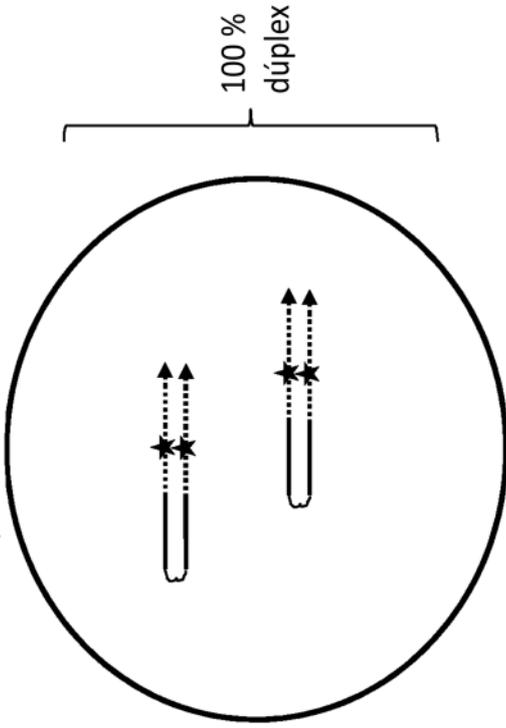


Figura 37A

Etapa 4: Agotar los cebadores diana enlazados (4-6 ciclos, o más)

Salida: 100 % de productos enlazados con identificación dúplex



Etapa 5: Cebado con cebadores enlazados universales

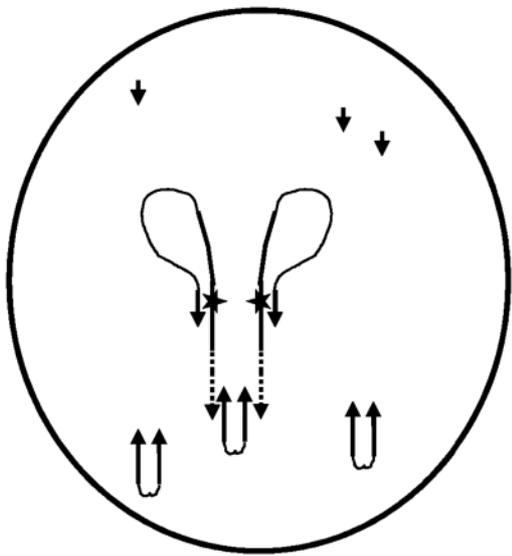


Figura 37B

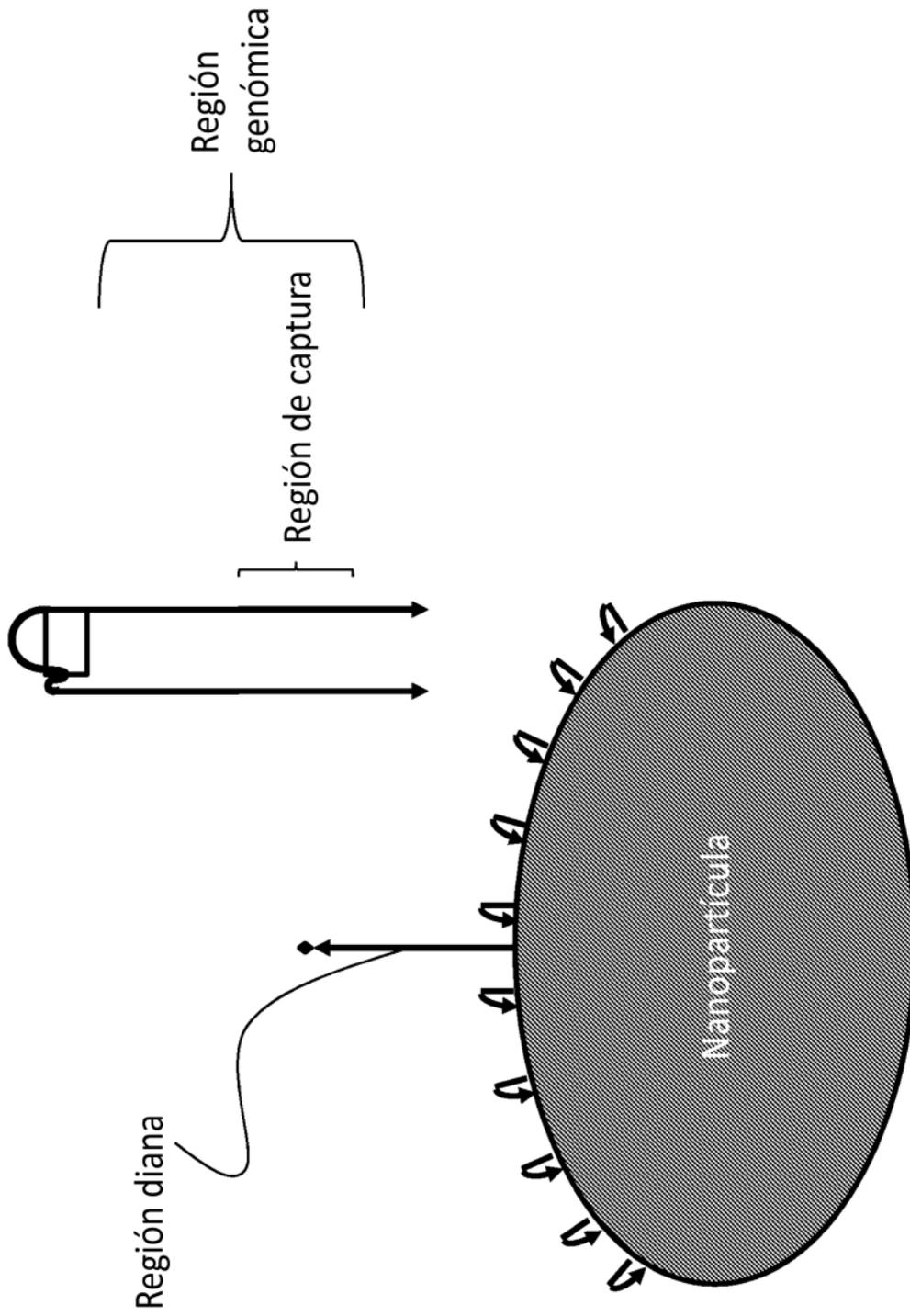


Figura 38

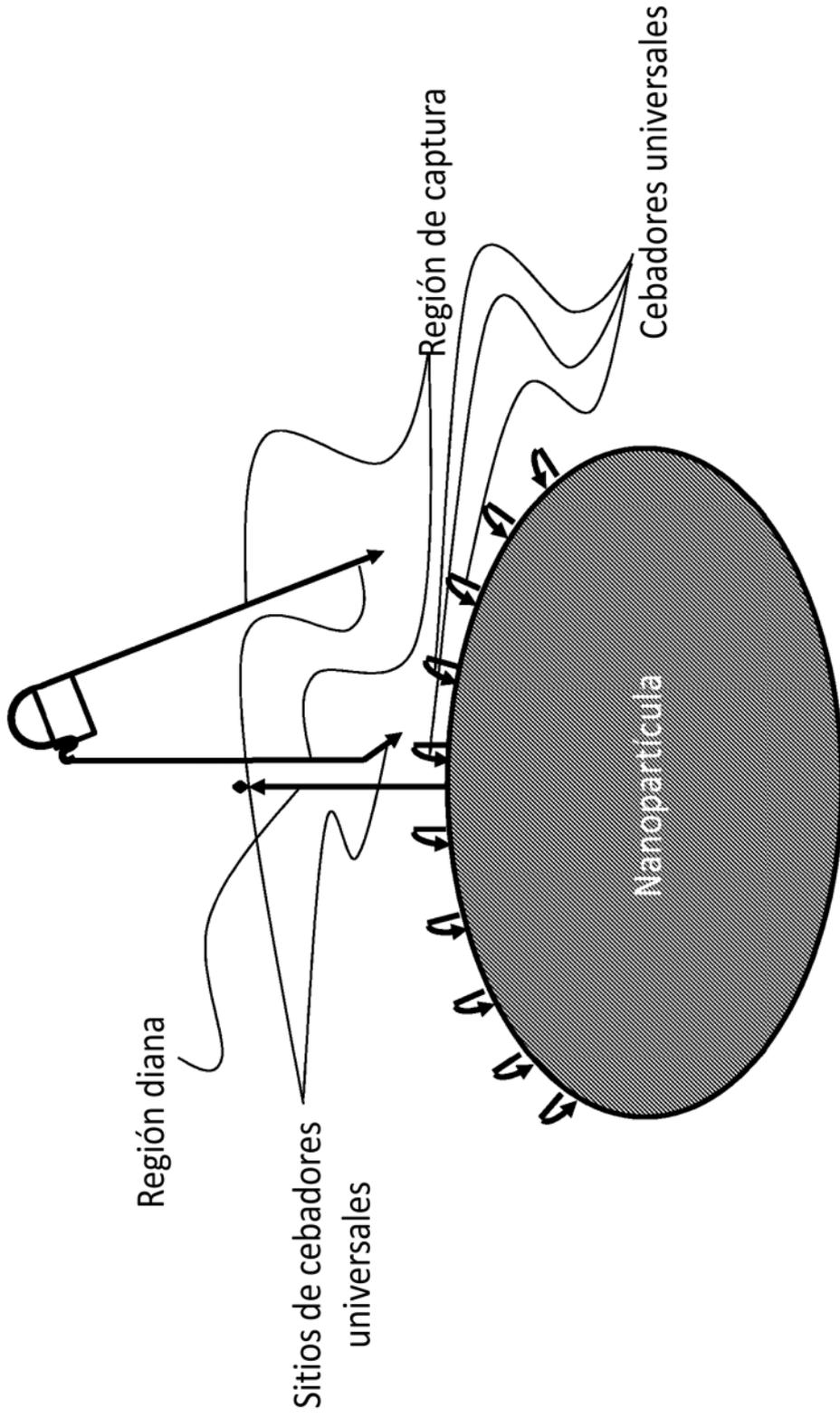


Figura 39

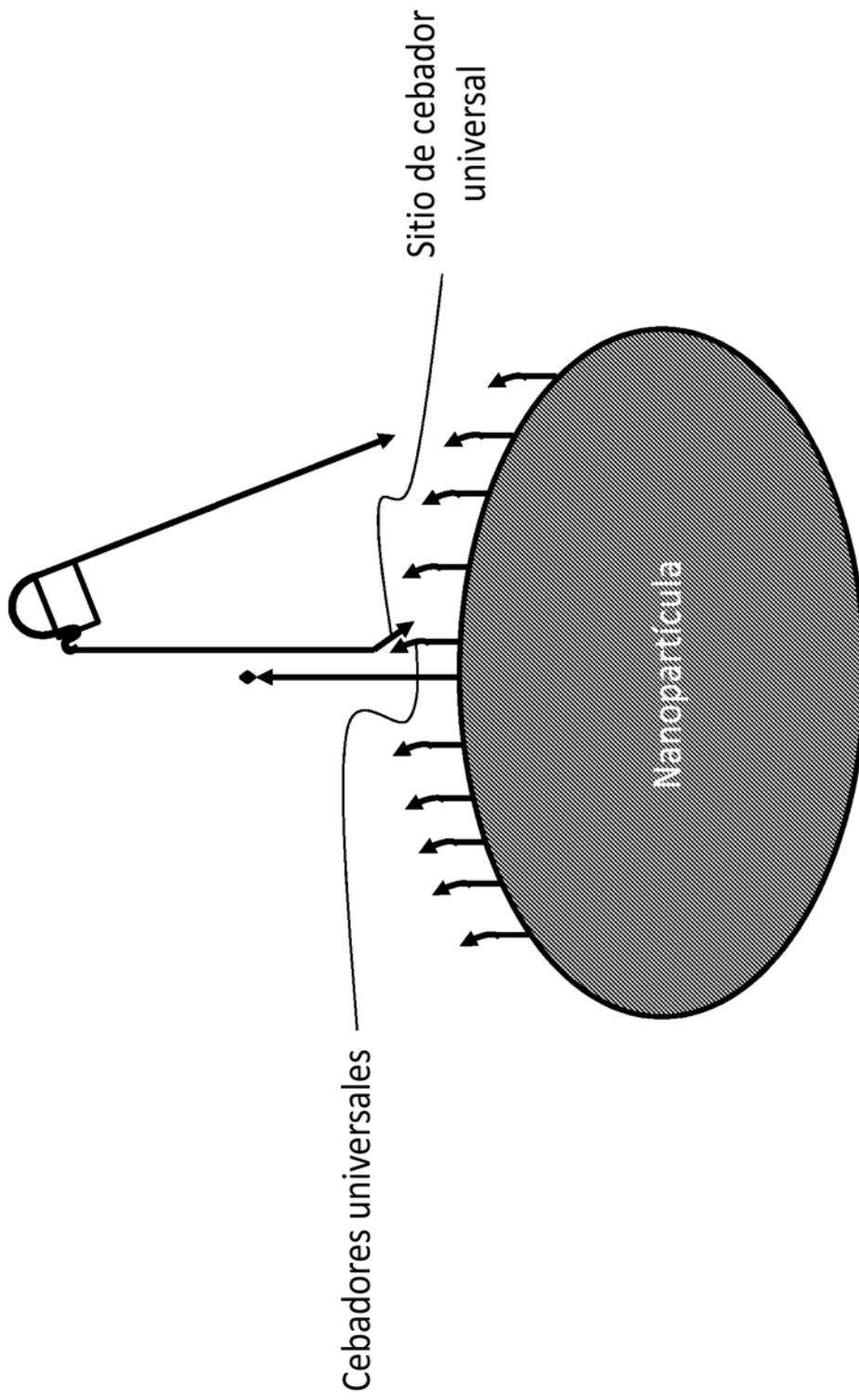


Figura 40

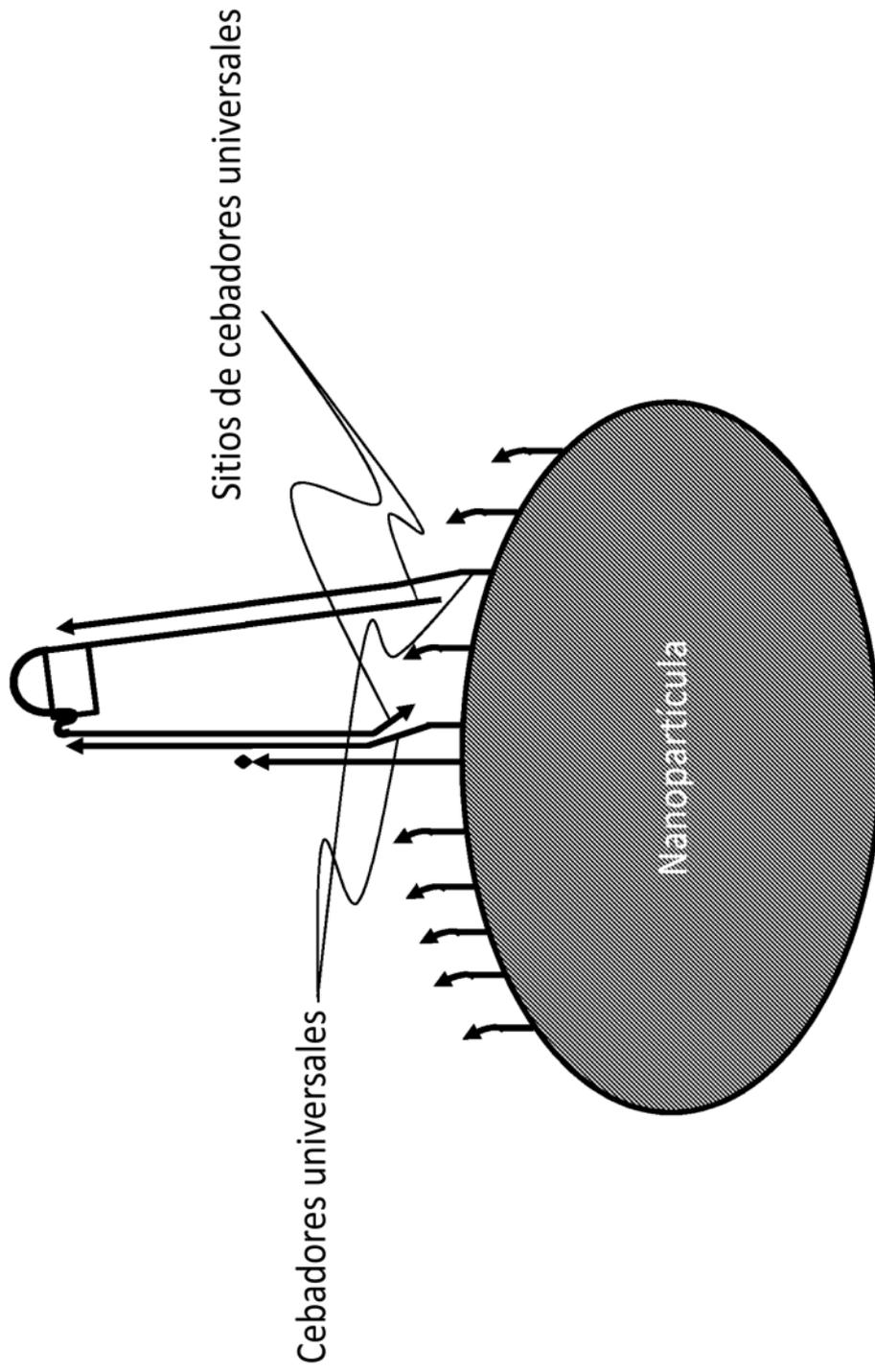


Figura 41

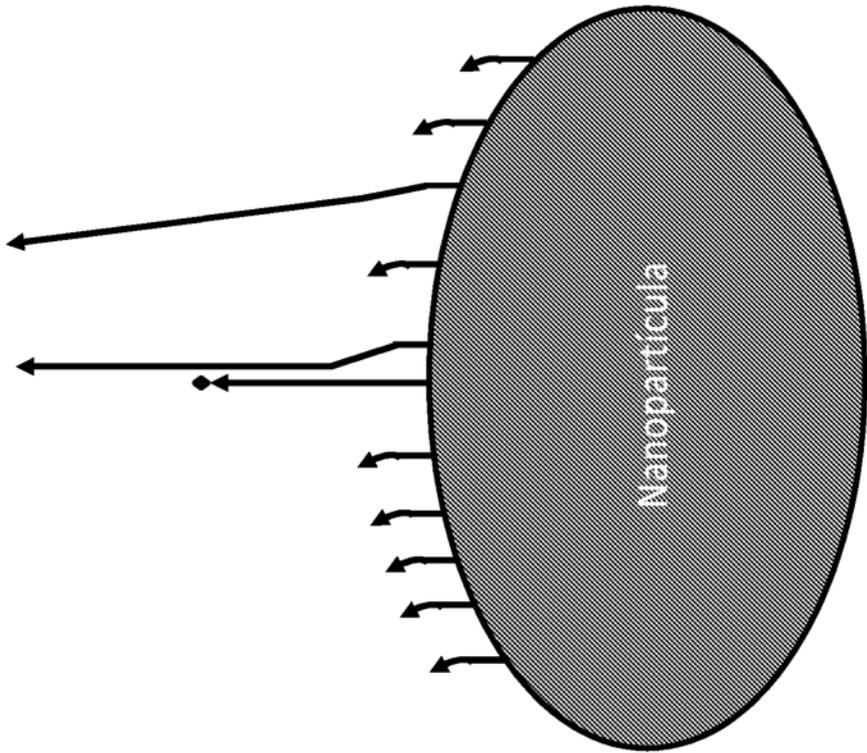


Figura 42

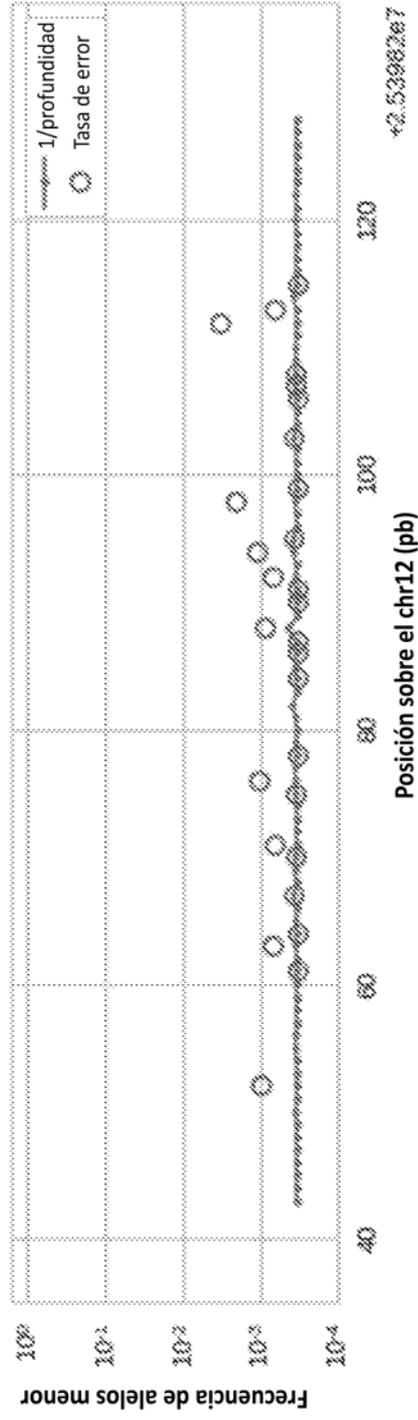


Figura 43