

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 020 408**

51 Int. Cl.:

C12Q 1/6874 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.02.2023** **PCT/US2023/061873**

87 Fecha y número de publicación internacional: **10.08.2023** **WO23150633**

96 Fecha de presentación y número de la solicitud europea: **02.02.2023** **E 23718537 (6)**

97 Fecha y número de publicación de la concesión europea: **09.04.2025** **EP 4466378**

54 Título: **Cebadores multifuncionales para lecturas de secuenciación por pares**

30 Prioridad:

02.02.2022 US 202263267466 P

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

22.05.2025

73 Titular/es:

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

72 Inventor/es:

KENNEDY, ANDREW

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 3 020 408 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

Cebadores multifuncionales para lecturas de secuenciación por pares

5 Antecedentes

Los métodos originales de secuenciación del ADN eran la síntesis dideoxídica de Sanger y la escisión química de Maxam-Gilbert. El método de Maxam-Gilbert se basa en la modificación química del ADN y la posterior escisión de la cadena principal del ADN en las zonas adyacentes a los nucleótidos modificados. La secuenciación de Sanger utiliza nucleótidos específicos de terminación de cadena (nucleótidos dideoxi) que carecen de un grupo 3'-OH, lo que da como resultado la terminación de la cadena de ADN en crecimiento en esa posición. Los ddNTPs se etiquetan radioactiva o fluorescentemente para su detección en geles de secuenciación o máquinas de secuenciación automática.

La necesidad de secuenciar grandes genomas con mayor rendimiento y menor coste impulsó el desarrollo de técnicas de secuenciación de nueva generación. Muchas técnicas de nueva generación distribuyen las moléculas individuales de ADN que se van a secuenciar en millones de perlas, pocillos o cámaras individuales o las unen a ubicaciones direccionables en un soporte plano. Las moléculas de ADN, a menudo tras la amplificación, se someten a reacciones de síntesis de ADN en las que se utilizan nucleótidos etiquetados, o a reacciones químicas basadas en la incorporación de un nucleótido concreto, pueden ser visualizadas o detectadas de otro modo.

Aunque los métodos de nueva generación generan una cantidad mucho mayor de secuencias a un coste menor que los métodos originales, suelen basarse en lecturas mucho más cortas (p. ej., de hasta aproximadamente 300-500 bases). Además, suelen tener un porcentaje de error superior al de los métodos originales.

El documento US20210189460 describía composiciones y métodos para determinar las secuencias de nucleótidos de secuencias de interés mediante secuenciación de extremos emparejados.

Sumario de la invención reivindicada

La invención proporciona un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana. El método comprende (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprende una región de unión a la diana 3' y un tallo-bucle 5' para formar una mezcla de reacción, en donde la región de unión a la diana se une al adaptador 3'; (b) ligar un extremo 5' libre del tallo-bucle al adaptador 3'; (c) realizar una reacción de extensión, en donde la región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y, opcionalmente, proporciona una lectura de secuenciación del ácido nucleico diana; (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle; (e) realizar una reacción de extensión, en donde el cebador de bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, y la primera cadena complementaria se desplaza del dúplex en formación con el ácido nucleico diana; (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la primera cadena complementaria desplazada; y g) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la primera cadena complementaria desplazada y proporciona una lectura de secuenciación de la primera cadena complementaria desplazada.

Opcionalmente, el método comprende además, antes de la etapa (a), poner en contacto el ácido nucleico diana monocatenario con un segundo cebador directo que comprende una región de unión a la diana, que se une a un sitio complementario en el adaptador 3'; realizar una reacción de secuenciación por extensión en donde la región de unión a la diana del segundo cebador directo inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y desplazar la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana. Opcionalmente, la etapa (c) es una reacción de secuenciación por extensión que proporciona una lectura de secuencia del ácido nucleico diana. Opcionalmente, la etapa (c) se realiza antes de la etapa (b).

La invención proporciona además un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana. El método comprende (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprende una región de unión a la diana 3' y un tallo-bucle 5' para formar una mezcla de reacción, en donde la región de unión a la diana se une al adaptador 3' y en donde la región de unión a la diana comprende un sitio escindible; (b) ligar un extremo 5' libre del tallo-bucle al adaptador 3'; (c) realizar una reacción de secuenciación por extensión, en donde la región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; escindir el sitio escindible, y realizar otra reacción de extensión en donde el resto de la región de unión a la diana después de la escisión inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana; (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle; (e) realizar una reacción de extensión en donde el cebador de

bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana; (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la cadena complementaria adicional desplazada; (g) realizar una reacción de secuenciación por extensión, en donde el cebador

5 inverso inicia la síntesis de una cadena complementaria de la cadena complementaria adicional desplazada y proporciona una lectura de secuenciación de la cadena complementaria adicional desplazada.

La invención proporciona además un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana. El método comprende (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo

10 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo en horquilla que comprende una región de unión a la diana 3' y un tallo-bucle 5' y un cebador directo lineal para formar una mezcla de reacción, en donde la región de unión a la diana del cebador directo de horquilla-tallo y el cebador directo lineal unen el adaptador 3' con el cebador de tallo-bucle entre el cebador lineal y el extremo 3'; (b) ligar un extremo 5' libre del tallo-bucle con el extremo 3' de la molécula de adaptador en el extremo 3' del ácido nucleico diana;

15 (c) realizar una reacción de secuenciación por extensión, en donde el cebador lineal inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y realizar otra reacción de extensión en donde la región de unión a la diana del cebador directo de tallo-bucle inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana; (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle; (e) realizar una reacción de extensión en

20 donde el cebador de bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana; (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la cadena complementaria adicional desplazada; y g) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la cadena complementaria adicional desplazada.

25 Opcionalmente, el cebador directo de tallo-bucle contiene un nucleótido bloqueado para la extensión y el método comprende además eliminar el bloqueo antes de extender el cebador directo de tallo-bucle.

La invención proporciona además un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana. El método incluye (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5'

30 por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprende una primera y una segunda regiones de unión a la diana 3' flanqueadas por una región intermedia para formar una mezcla de reacción, en donde la primera y la segunda regiones de unión a la diana se unen a sitios complementarios en la molécula de adaptador 3', estando los respectivos sitios en posición distal y proximal al extremo 3' del adaptador 3'; (b) realizar una reacción de extensión, en donde la primera región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y, opcionalmente, proporciona una lectura de secuenciación del ácido nucleico diana; (c) realizar una segunda reacción de extensión en donde la segunda región de unión a la diana inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria del dúplex en formación con el ácido nucleico diana; (d) poner en contacto la mezcla de

40 reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la primera cadena complementaria desplazada; y (e) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la primera cadena complementaria desplazada y proporciona una lectura de secuenciación de la primera cadena complementaria desplazada. Opcionalmente, el método comprende además, antes de la etapa (a), poner en contacto el ácido nucleico diana monocatenario con un segundo cebador directo que comprende una región de unión a la diana, que se une a un sitio complementario de la molécula de adaptador en el extremo 3' del ácido nucleico diana; realizar una reacción de secuenciación por extensión en donde la región de unión a la diana del segundo cebador directo inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y desplazar la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana. Opcionalmente, la región intermedia comprende un nucleótido invertido. Opcionalmente, la región intermedia comprende un enlace no nucleotídico. Opcionalmente, la segunda región de unión a la diana está bloqueada para la extensión y el método comprende además eliminar el bloqueo para la extensión antes de realizar la extensión iniciada por la segunda región de unión a la diana. Opcionalmente, la segunda región de unión a la diana está bloqueada para la extensión por un grupo fosfato terminal 3' y la eliminación del bloqueo comprende el tratamiento con una fosfatasa para eliminar el

50 fosfato. Opcionalmente, la segunda extensión se realiza con una polimerasa de desplazamiento de cadena. Opcionalmente, el adaptador 3' comprende además un marcador de muestra entre los sitios complementarios para la primera y la segunda regiones de unión a la diana.

En algunos métodos, el adaptador 5' o 3' o ambos comprenden además un marcador molecular. En algunos métodos,

60 el adaptador 5' o 3' o ambos comprenden además un marcador de muestra. En algunos métodos, el ácido nucleico monocatenario flanqueado por los adaptadores 5' y 3' se produce por amplificación en emulsión. Algunos métodos comprenden además romper la emulsión antes de realizar otras etapas descritas anteriormente. En algunos métodos, el soporte en el que se inmoviliza el ácido nucleico diana monocatenario se utilizó en la PCR en emulsión para generar el ácido nucleico diana monocatenario flanqueado por los adaptadores 5' y 3'. En algunos métodos, el soporte en el

65 que está inmovilizado el ácido nucleico diana monocatenario se fijó después de la PCR en emulsión. En algunos métodos, el ácido nucleico diana monocatenario inmovilizado en un soporte forma parte de una población clonal de

- dichos ácidos nucleicos diana inmovilizados en el mismo soporte. En algunos métodos, el soporte es una perla. En algunos métodos, el soporte es una región direccionable dentro de una matriz. En algunos métodos, la secuenciación es mediante secuenciación 454, de Ion Torrent o de Ultima. En algunos métodos, la secuenciación es de una sola molécula. En algunos métodos, la secuenciación es clonal. En algunos métodos, la longitud del ácido nucleico diana supera la longitud máxima de lectura del método de secuenciación. En algunos métodos, el ácido nucleico diana tiene aproximadamente 170 nucleótidos y la longitud máxima de lectura del método de secuenciación es de aproximadamente 150 nucleótidos del ácido nucleico diana. En algunos métodos, las lecturas de secuenciación tienen una región que no se solapa.
- 10 Algunos métodos comprenden además determinar una lectura de secuenciación compuesta a partir de las lecturas de secuenciación directa e inversa en donde cualquier posición de discordancia entre las lecturas de secuenciación directa e inversa se deja abierta o como nucleótidos alternativos que ocupan la posición en las lecturas de secuenciación directa e inversa.
- 15 La invención proporciona además un cebador que comprende una primera y una segunda regiones de unión a la diana, cada una con un extremo 3' libre separado por un nucleótido invertido.
- La invención proporciona además un kit que comprende un cebador con una primera y una segunda regiones de unión a la diana, cada una con un extremo 3' libre separado por un nucleótido invertido, primer y segundo adaptadores, o un adaptador único con regiones monocatenarias asimétricas dispuestas para flanquear cadenas individuales de ácido nucleico diana bicatenario con el primer y segundo adaptadores y un segundo cebador, en donde la primera y la segunda regiones de unión a la diana se unen al primer adaptador y el segundo cebador se une al segundo adaptador. Opcionalmente, uno de los extremos libres 3' está bloqueado con un grupo fosfato y el kit comprende además una fosfatasa para eliminar el fosfato.
- 25 El kit también puede incluir una polimerasa de desplazamiento de cadena.

Breve descripción de las figuras

- La figura 1 muestra una población clonal de un ácido nucleico diana monocatenario flanqueado por diferentes adaptadores monocatenarios con el adaptador 5' uniendo el ácido nucleico diana a un soporte. La parte derecha de la figura muestra un cebador multifuncional con una región de unión a la diana 3' y una región de tallo-bucle 5'.
- La figura 2 muestra el cebador multifuncional de la figura 1 hibridado con el adaptador en el extremo 3' del ácido nucleico diana a través de su región de unión a la diana.
- La figura 3 muestra la extensión iniciada por la región de unión a la diana de la figura 2 con el ácido nucleico diana sirviendo de molde, y la ligadura del extremo 5' del adaptador de tallo-bucle con el extremo 3' del adaptador 3' unido al ácido nucleico diana.
- La figura 4 muestra la extensión de la región de unión a la diana mencionada en la figura 3 que genera una lectura de secuenciación directa.
- La figura 5 muestra la extensión a la que se hace referencia en la figura 4 que ha generado una cadena complementaria del ácido nucleico diana (primera cadena complementaria), y un cebador de bucle que hibrida con la porción de bucle del cebador multifuncional.
- La figura 6 muestra la extensión del cebador de bucle que genera una segunda cadena complementaria y el desplazamiento de la cadena complementaria original a la que se hace referencia en la figura 5, que permanece unida al ácido nucleico diana por el cebador multifuncional y, por tanto, fijada al soporte.
- La figura 7 muestra un cebador de secuenciación inversa que hibrida con el complemento del adaptador en el extremo 5' del ácido nucleico diana en la primera cadena complementaria.
- La figura 8 muestra la extensión del cebador de secuenciación inversa que genera una lectura de secuenciación inversa.
- La figura 9 a la izquierda muestra un cebador multifuncional con una región de unión a la diana 3' y una región de tallo-bucle 5' como en la figura 1 pero aquí, la región de unión a la diana incluye un residuo de uracilo. A la derecha, la figura muestra un cebador multifuncional con una región de unión a la diana 3' y una región de tallo-bucle 5' bloqueada para la extensión por un grupo fosfato en el extremo 3'. También se muestra un cebador lineal que hibrida con el adaptador 3' cadena abajo (en dirección 3') del cebador multifuncional.
- La figura 10 muestra una población clonal de un ácido nucleico diana monocatenario flanqueado por diferentes adaptadores monocatenarios con el adaptador 5' uniendo el ácido nucleico diana a un soporte. El adaptador 3' incluye dos sitios de unión del cebador que flanquean un marcador de muestra opcional.

La figura 11 muestra un cebador multifuncional con dos regiones de unión a la diana 3' que hibridan con el adaptador 3' mostrado en la figura 10. Las dos regiones de unión a la diana 3' están separadas por una región intermedia que incluye una base invertida (dT).

- 5 La figura 12 muestra la extensión desde la región de unión a la diana 3' próxima al ácido nucleico diana que genera una cadena complementaria (primera cadena complementaria) y una lectura de secuenciación directa.

La figura 13 muestra el desbloqueo del extremo 3' de la región de unión a la diana 3' distal al ácido nucleico diana como preparación para su extensión.

- 10 La figura 14 muestra la extensión desde la región de unión a la diana 3' distal del ácido nucleico diana proporcionando inicialmente una lectura de secuenciación del marcador de muestra.

- 15 La figura 15 muestra la extensión completada desde la región de unión a la diana 3' distal del ácido nucleico diana que genera otra cadena complementaria y que desplaza la cadena complementaria original, que permanece unida al ácido nucleico diana por el cebador multifuncional y, de este modo, fijada al soporte.

- 20 La figura 16 muestra la hibridación de un cebador de secuenciación inversa con el complemento del adaptador en el extremo 5' de la diana, que ahora se encuentra en el extremo 3' de la cadena complementaria original desplazada. La extensión del cebador de secuenciación inversa con la cadena complementaria original desplazada que sirve de molde genera una lectura de secuenciación inversa.

Definiciones

- 25 Un "sujeto" se refiere a un animal, tal como una especie de mamífero (preferentemente humana) o aviar (p. ej., un ave), u otro organismo, tal como una planta. Más específicamente, un sujeto puede ser un vertebrado, p. ej., un mamífero, tal como un ratón, un primate, un simio o un humano. Entre los animales se incluyen los de granja, animales de deporte y mascotas. Un sujeto puede ser un individuo sano, un individuo que tiene síntomas o signos o se sospecha que tiene una enfermedad o una predisposición a la enfermedad, o un individuo que necesita terapia o se sospecha que la necesita.

- 30 Un ácido nucleico diana se refiere a un ácido nucleico cuya secuencia debe determinarse, al menos parcialmente. Un ácido nucleico diana puede formar parte de una población de ácidos nucleicos, que puede incluir otros ácidos nucleicos diana y ácidos nucleicos no diana presentes en la muestra. Los ácidos nucleicos de una muestra pueden denominarse colectivamente ácidos nucleicos de la muestra. Los ácidos nucleicos diana pueden ser ácidos nucleicos exentos de células o derivados de ácidos nucleicos exentos de células.

- 40 La secuenciación puede determinar variaciones genéticas. Una variación genética se refiere a un cambio en la secuencia de nucleótidos (variación nucleotídica), modificación o número de copias respecto a una secuencia de referencia, que puede ser, p. ej., un exón, gen, cromosoma o genoma completo que representa la secuencia normal, modificación, en su caso, y el número de copias de un organismo. Una variación genética puede incluir una o más variaciones de un solo nucleótido (SNV), inserciones, deleciones, repeticiones, pequeñas inserciones, pequeñas deleciones, pequeñas repeticiones, entrecruzamientos de variantes estructurales, repeticiones en tándem de longitud variable y/o secuencias flanqueantes, variantes de número de copia (CNV), transversiones, fusiones de genes y otras reordenaciones, así como modificaciones tales como la metilación, la acetilación o la hidroximetilación también son formas de variación genética. Una variación puede ser un cambio de base, inserción, delección, repetición, variación del número de copias, modificación, transversión, o cualquier combinación de los mismos.

- 50 Un marcador de cáncer es una variación genética asociada con la presencia o el riesgo de desarrollar un cáncer. Un marcador de cáncer puede indicar que un sujeto tiene cáncer o un mayor riesgo de desarrollar cáncer que un sujeto de la misma especie, de la misma edad y sexo, que no tenga el marcador de cáncer. Un marcador de cáncer puede o no ser causante de cáncer.

- 55 Los cuatro tipos de nucleótidos estándar se refieren a A, C, G, T para los desoxirribonucleótidos y A, C, T y U para los ribonucleótidos.

- 60 Dentro de una lectura de secuenciación, las expresiones "cadena arriba (en dirección 5')" y "cadena abajo (en dirección 3')" se utilizan para indicar secuencias relativamente más cercanas o más lejanas al punto de inicio de la secuenciación, normalmente un sitio de unión del cebador de secuenciación. Por ejemplo, si una lectura de secuenciación incluye un código de barras molecular cadena arriba (en dirección 5') y cadena abajo (en dirección 3'), el código de barras molecular cadena arriba (en dirección 5') está más cerca que el código de barras molecular cadena abajo (en dirección 3') del punto de inicio de la secuenciación.

- 65 Un cebador directo es un cebador que inicia la síntesis de la primera cadena a partir de un adaptador u otro molde, y un cebador inverso es un cebador que inicia la síntesis de la segunda cadena.

A menos que resulte evidente de otro modo por el contexto, la referencia a un ácido nucleico puede incluir ADN, o ARN o híbridos de los mismos. Las moléculas de ácido nucleico aisladas de la naturaleza suelen contener nucleótidos estándar, incluidas sus formas naturalmente modificadas, tales como la metilcitosina. Los oligonucleótidos sintéticos, tales como los adaptadores, también pueden formarse enteramente a partir de estos nucleótidos estándar, o pueden

El término "soporte" se refiere a cualquier artículo sólido o semisólido sobre el que pueden inmovilizarse reactivos tales como moléculas de ácido nucleico. Las moléculas de ácido nucleico pueden sintetizarse, unirse, ligarse, hibridarse o inmovilizarse de otro modo. Las moléculas de ácido nucleico pueden inmovilizarse en un sustrato por cualquier método, incluida la adsorción física, mediante la formación de enlaces iónicos o covalentes, o combinaciones de los mismos. Los soportes pueden ser planos, sustancialmente planos, o no planos. Un soporte puede ser bidimensional (p. ej., un sustrato 2D plano) o tridimensional. El soporte puede ser en forma de perlas, esferas, partículas, gránulos, un gel o una matriz porosa. Un soporte puede ser un componente de una celda de flujo y/o puede estar incluido dentro de un instrumento de secuenciación o adaptado para ser recibido por éste. Un soporte puede incluir un polímero, un vidrio o un material metálico, polímeros orgánicos tales como poliestireno, polietileno, polipropileno, polifluoroetileno, polietileno y poliacrilamida (p. ej., gel de poliacrilamida), así como copolímeros e injertos de los mismos. Un soporte puede comprender látex o dextrano. Un soporte puede ser inorgánico, tales como vidrio, sílice, oro, vidrio de poro controlado (CPG) o sílice de fase inversa. Algunos ejemplos de sustratos incluyen una membrana, un sustrato plano, una placa de microvaloración, una perla (p. ej., una perla magnética), un filtro, una tira reactiva y un portaobjetos. Un soporte puede ser un único artículo sólido o semisólido (p. ej., una única partícula), o una pluralidad de tales artículos (p. ej., una colección de partículas). Los sustratos pueden ser porosos o no porosos, y pueden tener características hinchables o no hinchables. Un soporte puede tener una forma para comprender uno o varios pocillos, depresiones u otros recipientes, vasos, rasgos distintivos o ubicaciones. Se puede configurar una pluralidad de soportes en una matriz en varias ubicaciones. Un soporte o los rasgos distintivos dentro de un soporte pueden ser direccionables (p. ej., para la administración robótica de reactivos o detección, por ejemplo, un sustrato puede estar en comunicación óptica y/o física con un detector o separado físicamente de un detector. Un soporte n puede colocarse dentro o sobre otro soporte (p. ej., dentro de un pocillo de un segundo soporte) para las reacciones de secuenciación y la detección.

Un cebador es un ácido nucleico capaz de hibridar con un sitio dentro de un molde de ácido nucleico e iniciar la extensión de una cadena complementaria del molde. Un cebador incluye una región de unión a la diana, que efectúa la hibridación con el molde. La región de unión a la diana suele tener un extremo 3' libre. Las longitudes de ejemplo de la región de unión a la diana son de al menos 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos, opcionalmente hasta 30, 35, 40, 45, 50 o más nucleótidos, incluyendo todas las permutaciones de los límites superior e inferior. La región de unión a la diana es perfectamente complementaria con el correspondiente sitio de la cadena molde con la que hibrida, o al menos lo suficientemente complementaria como para permitir la hibridación y la extensión. Un cebador también puede incluir un segmento 5' no complementario de la cadena molde, tal como una horquilla que no participa directamente en la hibridación con el molde o en la extensión desde la región de unión a la diana, sino que proporciona una funcionalidad adicional. Por ejemplo, el segmento 5' puede incluir un primer subsegmento y un segundo subsegmento unidos por un segmento intermedio, en donde el primer y segundo subsegmentos son complementarios en orientación opuesta, y en donde el segmento intermedio es capaz de doblarse sobre sí mismo para permitir que el primero y segundo subsegmentos se emparejen entre sí y formen un dúplex bicatenario. El primer y segundo subsegmentos del segmento 5' pueden estar situados en los extremos del segmento 5'. El segmento 5' de un cebador de tallo-bucle puede unirse al segmento 3' directamente o a través de una o más unidades de nucleobase intermedias. Las longitudes de ejemplo de los segmentos 5' de los cebadores de tallo-bucle son de al menos 12, 15, 20 o 25 unidades de nucleobase y, opcionalmente, de hasta 35, 40, 45, 50, 55, 60 o más unidades de nucleobase, incluidas todas las permutaciones de los límites superior e inferior. Un segmento 5' de un cebador también puede incluir otras funcionalidades, tales como una etiqueta para el etiquetado. Los cebadores pueden estar formados en su totalidad por los cuatro desoxirribonucleótidos estándar, o pueden incluir nucleótidos no naturales o enlaces no naturales entre nucleótidos, o enlaces no nucleotídicos como se describe adicionalmente más adelante.

Una reacción de extensión de cebador se refiere a la hibridación de un cebador con una molécula de cadena de ácido nucleico molde, seguido de la elongación del cebador para generar una cadena complementaria del molde. La extensión puede incorporar nucleótidos que se encuentran en ácidos nucleicos naturales o análogos, a veces etiquetados y/o de terminación de cadena a efectos de secuenciación. La extensión puede o no formar parte de una secuenciación por reacción de síntesis. Por secuenciación por síntesis se entiende una extensión en la que se detecta la identidad de los nucleótidos sucesivos añadidos en una extensión para determinar la secuencia. El cebado y la extensión de una cadena molde pueden o no ir precedidos de la desnaturalización de un molde bicatenario. El cebado y la extensión de una cadena molde pueden o no ir seguidos del cebado y la extensión de una cadena complementaria.

Una polimerasa es una enzima capaz de extender un cebador hibridado con un molde mediante la incorporación de

nucleótidos. Algunos ejemplos de polimerasas son la ADN polimerasa T7, ADN polimerasa del bacteriófago T4, PHL, ADN polimerasa 29 (phi29), polimerasa Taq, polimerasa Tth, polimerasa Tli, polimerasa Pfu, polimerasa Pwo, polimerasa VENT, polimerasa DEEPVENT, polimerasa EX-Taq, polimerasa LA-Taq, polimerasa Sso, polimerasa Poc, polimerasa Pab, polimerasa Mth, polimerasa ES4, polimerasa Tru, polimerasa Tac, polimerasa Tne, polimerasa Tma, polimerasa Tea, polimerasa Tih, polimerasa Tfi, polimerasas Platinum Taq, polimerasa Tbr, polimerasa Tfl, polimerasa Pftubo, polimerasa Pyrobest, polimerasa Pwo, polimerasa KOD, polimerasa Bst, polimerasa Sac, fragmento Klenow, polimerasa con actividad exonucleasa de 3' a 5'. Las polimerasas pueden tener o no actividad de desplazamiento de cadena para desplazar una cadena complementaria ya hibridada con un molde a medida que la polimerasa extiende una cadena naciente a partir de un cebador hibridado con el molde. Las ADN polimerasas T4 (NEB n.º M0203) y T7 (NEB n.º M0274) carecen de actividad de desplazamiento de cadena, mientras que phi29 (NEB n.º M0269) y la polimerasa SD (Ignatov, BioTechniques Volumen 57, Número 2, agosto de 2014, Páginas 81-87) tienen una gran capacidad para desplazar cadenas.

Los términos "dúplex" "hibridar" o "hibridación" y las variaciones gramaticales se refieren al emparejamiento específico de la secuencia de los ácidos nucleicos, por ejemplo entre un cebador y un molde, con una afinidad mayor que la existente entre cualquiera de los componentes de dicho emparejamiento y otros ácidos nucleicos que puedan estar presentes en una muestra, y en el caso de un cebador y un molde suficiente para que el cebador experimente una extensión dirigida por el molde.

Descripción detallada

I. General

La invención proporciona métodos para generar lecturas directas e inversas de un ácido nucleico diana monocatenario inmovilizado. La lectura directa puede obtenerse como una lectura directa de la diana monocatenaria desde su extremo 3' al extremo 5'. La lectura inversa puede obtenerse a partir de una cadena complementaria del ácido nucleico diana monocatenario desde el extremo 3' hasta el extremo 5' del complemento. El suministro de lecturas de secuencias directa e inversa del mismo ácido nucleico diana puede ser útil cuando la longitud del ácido nucleico diana supera la longitud de lectura de la tecnología de secuenciación. Las lecturas directa e inversa pueden combinarse después de convertir una de dichas lecturas en una secuencia de nucleótidos complementarios para proporcionar una longitud total de lectura mayor que cualquiera de las lecturas individuales. Cuando se sintetiza de forma independiente, la presencia de lecturas directa e inversa del mismo ácido nucleico diana también es útil para evitar o corregir errores, como indican las posiciones de discordancia entre las dos lecturas. Estas posiciones de discordancia pueden corregirse o evitarse en posteriores compilaciones de lecturas de secuenciación.

Por ejemplo, si una tecnología de secuenciación tiene una longitud de lectura máxima de 250 nucleótidos de un ácido nucleico diana (sin incluir la longitud de lectura adicional utilizada en las secuencias de los adaptadores flanqueantes) y un ácido nucleico diana tiene una longitud de 500 nucleótidos, entonces las lecturas directa e inversa pueden, conjuntamente, cubrir toda la longitud del ácido nucleico diana. Si la longitud de una lectura de secuenciación es de 150 nucleótidos de un ácido nucleico diana (de nuevo sin incluir las secuencias de los adaptadores flanqueantes) y el ácido nucleico diana es de 170 nucleótidos, entonces las lecturas directa e inversa pueden cubrir todo el ácido nucleico diana, además de proporcionar un solapamiento sustancial, que puede utilizarse para etiquetar errores en la secuencia leída en función de la discordancia.

Los métodos utilizan cebadores multifuncionales. Dicho cebador multifuncional sirve para iniciar síntesis separadas de la primera y segunda cadenas complementarias de un ácido nucleico diana monocatenario, y para anclar la primera cadena complementaria del ácido nucleico diana inmovilizado. La primera cadena complementaria sirve para proporcionar un sitio de unión del cebador y un molde para un producto de extensión para generar una lectura de secuenciación inversa. La segunda cadena complementaria sirve para desplazar a la primera cadena del dúplex en formación con el ácido nucleico diana. La primera o la segunda cadena complementaria pueden proporcionar una lectura de secuencia directa.

El material de partida para los métodos es un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' y en su extremo 3' por un adaptador 3'. Los adaptadores unidos al ácido nucleico diana monocatenario son monocatenarios. El ácido nucleico monocatenario diana se inmoviliza a través del adaptador 5' a un soporte. El soporte puede ser, por ejemplo, una perla o similar o una región direccionable de una matriz plana. El soporte puede ser el mismo que el utilizado en la síntesis o amplificación de un ácido nucleico diana. Como alternativa, el soporte utilizado para la síntesis o la amplificación, en su caso, puede retirarse antes de la secuenciación y un soporte unido a un ácido nucleico diana antes de la secuenciación. Los adaptadores, o a veces sus secuencias complementarias de los adaptadores monocatenarios, proporcionan sitios de unión del cebador. Los adaptadores 5' y 3' tienen sitios de unión del cebador diferentes, que pueden hibridar o formar dúplex con diferentes cebadores. La molécula diana de partida flanqueada por adaptadores puede formar parte de una población clonal de tales moléculas unidas al mismo soporte o a la misma región direccionable de un soporte, o puede ser una única molécula.

La combinación de uno o todos los ácidos nucleicos diana flanqueados por adaptadores, uno o más cebadores y reactivos para la extensión y secuenciación, tal como una polimerasa, nucleótidos y etiqueta(s) se denomina mezcla

de reacción. Los reactivos pueden mezclarse juntos inicialmente o introducirse en combinaciones o individualmente reflejando el orden utilizado como se describe más adelante.

El ácido nucleico diana monocatenario flanqueado por adaptadores monocatenarios puede generarse inicialmente como una molécula diana bicatenaria flanqueada por adaptadores bicatenarios. Por ejemplo, un ácido nucleico diana bicatenario puede ligarse a adaptadores bicatenarios, o los adaptadores pueden ligarse a un ácido nucleico diana en forma de etiquetas 5' a los cebadores utilizados en la amplificación. Antes o después de colocar los adaptadores, pueden amplificarse los ácidos nucleicos diana. Los adaptadores unidos a un ácido nucleico diana bicatenario pueden ser iguales o diferentes entre sí. La ligadura con un adaptador en forma de Y o equivalentes topológicos permite efectivamente que el mismo adaptador inicial, cuando se liga a un ácido nucleico diana bicatenario, genere cadenas componentes flanqueadas por diferentes adaptadores monocatenarios. Los adaptadores también pueden incluir códigos de barras moleculares y/o de muestra. Los ácidos nucleicos diana bicatenarios flanqueados por adaptadores bicatenarios pueden convertirse en monocatenarios por desnaturalización, amplificación o extensión, entre otros métodos. Los ácidos nucleicos diana bicatenarios o monocatenarios pueden unirse a un soporte. La unión puede realizarse tras la síntesis y, opcionalmente, desnaturalización de un ácido nucleico diana bicatenario o los ácidos nucleicos diana pueden sintetizarse nucleótido a nucleótido sobre un soporte o una matriz plana.

La amplificación en emulsión es una técnica que permite obtener un material de partida adecuado para los presentes métodos. La amplificación en emulsión ya constituye una etapa previa en varios protocolos de secuenciación, incluido el 454, Ion Torrent y Ultima. La amplificación en emulsión permite amplificar moléculas diana individuales unidas a adaptadores dentro de compartimentos separados formados por una emulsión. Dentro de cada pocillo se forma una población clonal de un ácido nucleico diana flanqueado por adaptadores 5' y 3' inmovilizados en una perla. Diferentes compartimentos de la emulsión contienen perlas con poblaciones clonales de diferentes ácidos nucleicos diana. Tras la amplificación, la emulsión puede romperse mezclando las perlas y sus poblaciones clonales de ácido nucleico diana unidas. Las etapas posteriores de hibridado y extensión del cebador pueden realizarse en las perlas mezcladas en solución o después de transferir las perlas a una celda de flujo de secuenciación.

Un método de la invención emplea un cebador directo que comprende una región de unión a la diana 3' y un tallo-bucle 5', y un cebador inverso lineal. La región de unión a la diana 3' está configurada para unirse a un adaptador 3' que flanquea un ácido nucleico diana. El cebador inverso está configurado para hibridar con un complemento (es decir, secuencia complementaria) del adaptador 5'. La longitud de la secuencia complementaria en los cebadores necesaria para dicha hibridación es comparable a la de los cebadores convencionales (p. ej., 10-25 nucleótidos contiguos). Normalmente, los cebadores se unen a sitios de unión del cebador complementarios dentro de un adaptador, pero sin ocupar toda la longitud del adaptador.

Un ácido nucleico diana flanqueado por adaptadores 5' y 3' se pone en contacto con un cebador directo como se ha descrito anteriormente en condiciones de hibridación. Para facilitar la ilustración, el método se describe para un único ácido nucleico diana, pero puede realizarse en paralelo con una población de ácidos nucleicos diana (p. ej., al menos 10, 100, 1000 o 1.000.000) generalmente flanqueados por los mismos adaptadores 5' y 3' (con la posible excepción de los códigos de barras moleculares y/o de muestra). El cebador directo hibrida con el adaptador 3' a través de su región de unión a la diana. El extremo 3' del cebador se sitúa a continuación para iniciar la extensión de una primera cadena complementaria de la molécula de ácido nucleico diana. Antes o después de iniciar dicha extensión, el extremo 5' del cebador de tallo-bucle se liga al extremo 3' del adaptador 3'. El extremo 3' del cebador directo se extiende con una polimerasa, sirviendo el ácido nucleico diana de molde y se sintetiza una primera cadena complementaria del ácido nucleico diana. La extensión puede ser una reacción de secuenciación por síntesis en la cual se detecta sucesivamente la identidad de los sucesivos tipos de nucleótidos añadidos al cebador directo para generar una secuencia de nucleótidos que constituye una lectura de secuenciación directa del ácido nucleico diana. La extensión también puede ser sin secuenciación, la cual puede denominarse extensión oscura. Si se trata de una extensión oscura, la secuencia puede determinarse por separado mediante una tecnología que no sea la secuenciación por síntesis. La extensión puede ser en parte secuenciación por síntesis y en parte extensión oscura, en cuyo caso la parte del ácido nucleico diana correspondiente a la extensión oscura puede secuenciarse por separado. Como alternativa, la secuencia puede determinarse llevando a cabo una reacción separada de secuenciación por síntesis del ácido nucleico diana, como se describe más adelante.

Una vez completada la síntesis de la primera cadena complementaria, ya sea mediante una reacción de secuenciación por síntesis y/o secuenciación oscura, se lleva a cabo una nueva extensión cebada por un cebador de bucle que hibrida con la región de bucle del cebador directo de tallo-bucle. Este cebador puede añadirse a la mezcla de reacción después de completar la primera extensión o puede haberse añadido previamente, en cuyo caso se puede bloquear el extremo 3' hasta que se vaya a realizar la extensión a partir del cebador de bucle. El bloqueo puede ser con un grupo fosfato y la eliminación con una fosfatasa. El cebador de bucle se adhiere al bucle del cebador directo e inicia la extensión con el ácido nucleico diana que sirve de nuevo de molde para generar una segunda cadena complementaria. Esta extensión puede ser una extensión oscura, que sirve para desplazar la cadena complementaria sintetizada previamente del ácido nucleico diana. Como alternativa, por ejemplo, si la extensión previa para generar una cadena complementaria es una extensión oscura, la generación de la segunda cadena complementaria puede ser una reacción de secuenciación por síntesis para generar una lectura de secuenciación directa. Esta extensión se realiza preferentemente con una polimerasa con actividad de desplazamiento de cadena. Como alternativa, el

desplazamiento puede efectuarse por otros medios, tales como la desnaturalización. La cadena complementaria desplazada permanece unida covalentemente al ácido nucleico diana por el cebador directo de tallo-bucle.

La primera cadena complementaria desplazada tiene en su extremo 3' una secuencia complementaria al adaptador 5'. Esta cadena complementaria proporciona un sitio de unión para un cebador inverso. El cebador inverso puede suministrarse una vez finalizada la síntesis de la cadena complementaria adicional y el desplazamiento de la cadena complementaria original, o en una fase anterior del proceso. El cebador inverso después de la hibridación con el complemento del adaptador 5', puede iniciar la extensión con la primera cadena complementaria (es decir, generada a partir del extremo 3' del cebador de tallo-bucle) que sirve de molde. Esta extensión puede ser una etapa de secuenciación por síntesis que genere una secuencia de nucleótidos que constituya una lectura de secuencia inversa. Otra posibilidad es que se trate de una extensión oscura y que la secuencia se determine posteriormente mediante una tecnología distinta de la secuenciación por síntesis.

El método descrito anteriormente genera la lectura de secuenciación inversa mediante la lectura de una cadena complementaria del ácido nucleico diana generada por la lectura de secuenciación directa. Por lo tanto, los errores presentes en la lectura de secuenciación directa también estarán presentes en las posiciones correspondientes de la lectura de secuenciación inversa. El método puede modificarse para generar lecturas de secuenciación directa e inversa independientes. Una de estas modificaciones consiste en utilizar un segundo cebador directo, normalmente antes de utilizar el cebador directo de tallo-bucle. El segundo cebador suele ser un cebador lineal que tiene una región de unión a la diana que forma un dúplex en un sitio dentro del adaptador 3', igual o diferente al dúplex formado por el cebador directo de tallo-bucle. El segundo cebador directo puede iniciar la síntesis de una cadena complementaria con el ácido nucleico diana como molde que genera una lectura de secuenciación del ácido nucleico diana. A continuación, esta cadena complementaria se desplaza mediante desnaturalización o digestión o mediante el uso de una polimerasa con actividad de desplazamiento de cadena para la siguiente extensión. A continuación, el método procede como se ha descrito anteriormente utilizando un cebador directo de tallo-bucle para iniciar la síntesis de una cadena complementaria del ácido nucleico diana. Pero si ya se ha determinado una lectura directa del ácido nucleico diana, esta extensión puede ser una extensión oscura. A continuación se produce la ligadura de la horquilla, la hibridación de un cebador de bucle, la extensión de otra cadena complementaria, el desplazamiento de la cadena hibridando un cebador inverso y la extensión para determinar una lectura de secuenciación de la cadena inversa y, a continuación, proceder como se ha descrito anteriormente.

Otra variación para obtener lecturas de secuenciación directa e inversa independientes utiliza el cebador directo de tallo-bucle, cebador de bucle y cebador inverso como se ha descrito anteriormente. Sin embargo, la extensión inicial del cebador de tallo-bucle para generar una primera cadena complementaria es una extensión oscura. La extensión posterior del cebador de bucle hibridado con el cebador de tallo-bucle de una segunda cadena complementaria del ácido nucleico diana, así como el desplazamiento de la primera cadena complementaria, es ahora una reacción de secuenciación por síntesis que proporciona una lectura de secuenciación directa.

Otra variación utiliza el cebador directo de tallo-bucle, cebador de bucle y cebador inverso como se ha descrito anteriormente, pero el cebador directo de tallo-bucle ahora contiene un nucleótido escindible dentro de su región de unión a la diana (figura 9 izquierda). Un ejemplo de nucleótido escindible es el uracilo o la desoxiuridina, que puede escindirse por digestión con ARNasa H. Se lleva a cabo una extensión cebada desde el extremo 3' del cebador directo de tallo-bucle, como se ha descrito anteriormente, para generar una cadena complementaria del ácido nucleico diana y una lectura de secuenciación directa. El nucleótido escindible se escinde ahora, y la cadena complementaria que se acaba de sintetizar y un trozo del cebador directo de tallo-bucle que queda después de la escisión se disocian del ácido nucleico diana. A continuación, se realiza otra extensión a partir de un extremo 3' recién generado del cebador directo de tallo-bucle. Esta extensión genera otra cadena complementaria del ácido nucleico diana. Si se realiza con una polimerasa con actividad de desplazamiento de cadena, esta extensión puede provocar el desplazamiento del extremo complementario generado antes de la escisión. La extensión para generar la cadena complementaria adicional puede ser una extensión oscura porque la lectura de secuenciación directa ya ha sido determinada. Después de esto, el método procede como se describió originalmente con la extensión del cebador de bucle hibridado con la región de bucle del cebador directo de tallo-bucle para generar otra cadena complementaria que desplaza a la cadena complementaria sintetizada previamente, lo que proporciona un sitio de unión para el cebador inverso. El cebador inverso inicia la síntesis del complemento de la cadena complementaria desplazada, lo que permite determinar una lectura de secuenciación inversa.

En otra variación, el cebador directo de tallo-bucle se utiliza junto con un cebador directo lineal diferente (figura 9 derecha). El cebador directo lineal y la región de unión a la diana del cebador directo de tallo-bucle hibridan en un sitio adyacente o próximo, de modo que el cebador directo de tallo-bucle se encuentre entre el cebador lineal y el extremo 3' del adaptador 3'. Juntos, el cebador directo lineal y el cebador directo de tallo-bucle pueden considerarse como un cebador dividido topológicamente análogo al cebador directo de tallo-bucle descrito anteriormente con un sitio escindible después de que se haya producido la escisión. El cebador directo lineal inicia la extensión de una primera cadena complementaria con el ácido nucleico diana como molde. Esta extensión puede ser una extensión de secuenciación por síntesis para proporcionar una lectura de secuenciación directa. A continuación, la extensión a partir del cebador directo de tallo-bucle procede de forma análoga a la extensión a partir del extremo del cebador directo de tallo-bucle generado por escisión, tal como se ha descrito anteriormente. Las etapas restantes son también las mismas

que para el cebador directo de tallo-bucle con un nucleótido escindible.

Un método alternativo para determinar lecturas de secuenciación directa e inversa de un ácido nucleico diana monocatenario flanqueado por adaptadores 5' y 3' como se ha descrito anteriormente e inmovilizado a través del adaptador 5', utiliza un cebador directo que comprende una primera y una segunda regiones de unión a la diana 3' que flanquean una región intermedia en lugar del cebador directo de tallo-bucle. La región intermedia contiene un enlace nucleotídico no estándar o un enlace no nucleotídico para permitir la unión de dos regiones de unión a la diana 3' en el mismo cebador. Un nucleótido invertido es un ejemplo de enlace adecuado. La primera y la segunda regiones de unión a la diana se unen a sitios complementarios en la molécula de adaptador 3', estando los respectivos sitios en posición distal y proximal al extremo 3' del adaptador 3'. En otras palabras, cuando se hibrida, la segunda región de unión a la diana se encuentra entre la primera región de unión a la diana y el extremo 3' del adaptador 3'. Las dos regiones de unión a la diana 3' proporcionan dos puntos separados para iniciar la extensión de una primera cadena complementaria desde dentro del adaptador 3'. La extensión desde el primer sitio de unión a la diana es análoga a la extensión desde el extremo 3' del cebador directo de tallo-bucle en el método descrito anteriormente y la extensión desde el segundo sitio de unión a la diana es análoga a la extensión desde el cebador de bucle hibridado con el cebador directo de tallo-bucle en el método descrito anteriormente. La segunda región de unión a la diana puede bloquearse para la extensión, tal como por la presencia de un grupo fosfato, para evitar que inicie la extensión hasta que se haya completado la extensión desde la primera región de unión a la diana. La extensión desde la primera región de unión a la diana genera una primera cadena complementaria del ácido nucleico diana. La extensión puede ser una reacción de secuenciación por síntesis para generar una lectura de secuenciación directa o una extensión oscura. Si es una extensión oscura, la secuencia puede determinarse mediante un método de secuenciación distinto de la secuenciación por síntesis o la lectura de secuenciación directa puede determinarse a partir de una extensión diferente, como se describirá más adelante. La extensión también puede ser en parte secuenciación por síntesis y en parte extensión oscura. Tras la generación de la primera cadena complementaria a partir de la primera síntesis de unión a la diana, el segundo sitio de unión a la diana se desbloquea, si es necesario, y se utiliza para iniciar la extensión de una segunda cadena complementaria con el ácido nucleico diana que sirve de molde. Esta extensión puede ser una extensión de secuenciación por síntesis para generar una lectura de secuenciación directa o una extensión oscura. La extensión se realiza preferentemente con una polimerasa con actividad de desplazamiento de cadena para desplazar la cadena complementaria sintetizada originalmente. El desplazamiento puede ser como alternativa el resultado de la desnaturalización. Después del desplazamiento, la primera cadena complementaria permanece unida covalentemente al cebador directo, que también está unido covalentemente a la segunda cadena complementaria, que a su vez forma un dúplex con la molécula de ácido nucleico diana, que se inmoviliza a través de su adaptador 5'. Por lo tanto, la primera cadena complementaria desplazada permanece anclada al ácido nucleico diana inmovilizado. Se suministra un cebador inverso que hibrida con una secuencia complementaria del adaptador 5' presente en el extremo 3' de la primera cadena complementaria desplazada. El cebador hibridado puede entonces iniciar la síntesis de una cadena complementaria de la primera cadena complementaria desplazada. Esta extensión puede ser una extensión de secuenciación por síntesis o la secuencia puede determinarse por separado una vez finalizada la extensión. En cualquier caso, se genera una lectura de secuencia inversa.

Como se ha descrito para el método del cebador de tallo-bucle, el método descrito puede modificarse para generar lecturas de secuenciación directa e inversa independientes. Por ejemplo, se puede realizar una extensión inicial con un segundo cebador directo, normalmente un cebador lineal, complementario a un sitio en el adaptador 3'. Esta extensión puede ser una reacción de secuenciación por síntesis para generar una lectura de secuenciación directa o se puede realizar la secuenciación después de la extensión para generar una lectura de secuenciación directa. Después de esta extensión, el cebador directo hibrida con la primera y la segunda regiones de unión a la diana 3'. El primer sitio de unión a la diana se utiliza para iniciar la síntesis de una cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria extendida del segundo cebador directo. Esta extensión normalmente es oscura. El segundo sitio de unión a la diana se utiliza para iniciar la síntesis de otra cadena complementaria para desplazar la cadena complementaria extendida desde la primera región de unión a la diana 3'. Esta extensión también es normalmente oscura. Un cebador inverso hibrida con el complemento del adaptador 5' ahora en el extremo 3' de la cadena complementaria desplazada y se procede a la secuenciación de la cadena inversa como se ha descrito anteriormente.

La práctica de los métodos anteriores genera una lectura de secuenciación directa y una lectura de secuencia inversa de uno o más ácidos nucleicos diana. Los tipos de nucleótidos de las lecturas de secuenciación inversa pueden convertirse en sus complementos para mantener la coherencia con la lectura de secuenciación directa o viceversa. Las lecturas de secuenciación pueden alinearse y compararse por pares, con lecturas de secuenciación de otro ácido nucleico diana o con una secuencia de referencia conocida. La alineación por pares de las lecturas de secuencia directa e inversa puede generar una lectura de secuenciación compuesta más larga que la lectura de secuenciación directa o inversa. En las posiciones de la alineación por pares en las que hay discordancia entre las lecturas de secuenciación directa e inversa, la discordancia puede indicarse no asignando un nucleótido en esa posición de la secuencia compuesta (es decir, dejándola abierta) o asignando una posición de ambigüedad en donde puede estar presente cualquiera de los nucleótidos presentes en las lecturas de las secuencias directa e inversa. La posición omitida o ambigua puede resolverse mediante alineación con lecturas de secuenciación de otros ácidos nucleicos diana de secuencia al menos solapada en los que la misma posición no esté sujeta a la misma omisión o ambigüedad. Las lecturas de secuenciación directa e inversa individuales o las lecturas de secuenciación compuestas procedentes

de la alineación por pares de lecturas de secuenciación directa e inversa del mismo ácido nucleico diana pueden agruparse en familias de acuerdo con la molécula de origen de la muestra y analizarse posteriormente como se describe a continuación.

5 II. Códigos de barras moleculares y de muestra y adaptadores

Tal como se ha mencionado anteriormente, los adaptadores pueden incluir códigos de barras de muestra y/o moleculares. Un código de barras es un ácido nucleico corto (p. ej., de menos de 500, 100, 50, 20, 15, 10 o 5 nucleótidos de longitud), utilizado para etiquetar moléculas de ácido nucleico para distinguir ácidos nucleicos de diferentes muestras (un código de barras de muestra), o diferentes moléculas de ácido nucleico en la misma muestra (un código de barras molecular) o el mismo código de barras puede utilizarse para distinguir tanto muestras como moléculas dentro de las muestras.

Los códigos de barras normalmente se proporcionan como conjuntos de múltiples códigos de barras individuales diferentes para distinguir muestras y moléculas, o ambas cosas. Es decir, diferentes muestras pueden recibir diferentes códigos de barras de muestra de un conjunto de códigos de barras de muestra, y diferentes moléculas dentro de una muestra reciben diferentes códigos de barras moleculares de un conjunto de códigos de barras moleculares. Los códigos de barras pueden ser monocatenarios, bicatenarios o tienen componentes tanto monocatenarios como bicatenarios. Los códigos de barras pueden tener la misma o distinta longitud dentro de un conjunto. Los códigos de barras pueden ser aleatorios, secuencias no aleatorias o semialeatorias en las que al menos una posición se selecciona aleatoriamente y al menos una no. Los códigos de barras pueden sintetizarse conjuntamente con la agrupación de nucleótidos en posiciones aleatorias, o individualmente. Algunos conjuntos de códigos de barras tienen secuencias seleccionadas de tal manera que hay una distancia de Hamming de al menos 2, 3, 4 o 5 nucleótidos entre cada código de barras de un conjunto. Los códigos de barras también pueden seleccionarse para evitar secuencias que hibriden entre sí o con otras moléculas en una reacción, para evitar secuencias sujetas a errores de secuenciación, o secuencias sujetas a confusión con secuencias de otros códigos de barras. Los códigos de barras como componentes de adaptadores o colas de cebadores de amplificación pueden unirse a uno o ambos extremos de los ácidos nucleicos que se van a etiquetar.

Los códigos de barras de las muestras pueden descodificarse para revelar la muestra de origen. Los códigos de barras de muestra permiten la agrupación y el procesamiento en paralelo de múltiples muestras una vez unidos los códigos de barras. El número de códigos de barras de una muestra diferente dentro de un conjunto suele ser suficiente para que cada muestra diferente esté asociada a un código de barras de muestra diferente o a una combinación de códigos de barras. Como alternativa, las muestras pueden dividirse en subconjuntos, de forma que las muestras de un subconjunto reciban el mismo código de barras de muestra y las muestras de subconjuntos diferentes reciban códigos de barras de muestra diferentes.

Los códigos de barras moleculares se utilizan para rastrear moléculas originales dentro de la misma muestra. Pueden descodificarse para revelar copias de amplificación o lecturas de secuenciación de la misma molécula original. El número de códigos de barras moleculares dentro de un conjunto o el número de combinaciones por pares dentro de un conjunto si las moléculas de la muestra están marcadas con códigos de barras moleculares por ambos extremos puede ser suficiente para que exista una alta probabilidad (p. ej., al menos el 80, 90, 95 o 99 % de probabilidad) de que sustancialmente todas las moléculas originales de la muestra que completen la ligadura con un adaptador o par de adaptadores (p. ej., al menos el 75 %, 90 %, 95 % o 99 %) reciban un código de barras molecular diferente o una combinación diferente de códigos de barras moleculares (código de barras único). Como alternativa, el número de códigos de barras moleculares o de combinaciones de códigos de barras moleculares por pares puede ser sustancialmente inferior al número de moléculas de una muestra, p. ej., una proporción de códigos de barras moleculares diferentes o una combinación por pares de códigos de barras moleculares con respecto a las moléculas de las muestras inferior a $1:10^3$, $1:10^4$, $1:10^5$, $1:10^6$, $1:10^7$, $1:10^8$, $1:10^9$, $1:10^{10}$, $1:10^{11}$ o $1:10^{12}$ código de barras no único). En este caso, múltiples moléculas dentro de la misma muestra reciben el mismo código de barras molecular o combinación de códigos de barras moleculares. Sin embargo, los productos de amplificación de la misma molécula original o sus lecturas de secuenciación aún pueden distinguirse utilizando una combinación de los códigos de barras moleculares y la información de las lecturas de secuenciación, tales como los puntos de inicio y/o parada o la longitud de las lecturas de secuenciación. Normalmente se utilizan suficientes códigos de barras moleculares diferentes o combinaciones de códigos de barras moleculares de tal forma que exista una alta probabilidad (p. ej., al menos el 90 %, al menos el 95 %, al menos el 98 %, al menos el 99 %, al menos el 99,9 % o al menos el 99,99 %) de que todos los ácidos nucleicos que corresponden a una región genómica particular definida por los mismos puntos de inicio y/o de parada lleven un código de barras molecular diferente. Generalmente, la asignación de códigos de barras moleculares únicos o no únicos en las reacciones sigue los métodos y sistemas descritos en las solicitudes de patente de Estados Unidos 20010053519, 20030152490, 20110160078, y la patente de Estados Unidos n.º 6.582.908 y la patente de Estados Unidos n.º 7.537.898.

En algunos casos, el número de códigos de barras moleculares diferentes es de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10.000, 50.000, 100.000, 500.000, 1.000.000, 10.000.000, 50.000.000 o 1.000.000.000. En otros casos, el número de códigos de barras moleculares diferentes es inferior a 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10.000, 50.000, 100.000, 500.000, 1.000.000, 10.000.000, 50.000.000 o 1.000.000.000 de códigos

de barras moleculares por muestra genómica. El número de códigos de barras moleculares diferentes en un conjunto depende de si se utiliza un código de barras único o no único y de si los códigos de barras moleculares se utilizan para etiquetar moléculas de muestra de ácido nucleico individualmente o en combinaciones de pares. En igualdad de condiciones, se necesitan más códigos de barras moleculares diferentes para el etiquetado único que para el no único.

- 5 También se necesitan más códigos de barras moleculares diferentes para el etiquetado con códigos de barras moleculares individuales por ácido nucleico de muestra que en combinaciones por pares, porque el número de combinaciones es el cuadrado del número de etiquetados individuales.

- 10 El número de códigos de barras moleculares diferentes necesarios para el etiquetado único de moléculas de ácidos nucleicos depende de cuántas moléculas de ácido nucleico originales haya en la muestra o parte de la misma que se esté analizando. Esto, a su vez, depende de factores como el número total de equivalentes de genoma haploide de la muestra, la media y la varianza del tamaño de las moléculas de ácido nucleico, y la eficacia de ligadura de los adaptadores, incluidos los códigos de barras.

- 15 Para códigos de barras no únicos, el número de combinaciones de códigos de barras moleculares (cuadrado del número de códigos de barras moleculares diferentes) es a veces al menos cualquiera de 64, 100, 400, 900, 1400, 2500, 5625, 10.000, 14.400, 22.500 o 40.000 y no más de cualquiera de 90.000, 40.000, 22.500, 14.400 o 10.000. Por ejemplo, el número de combinaciones de códigos de barras puede estar entre 64 y 90.000, entre 400 y 22.500, 400 y 14.400 o entre 900 y 14.400. El número de combinaciones diferentes de códigos de barras moleculares (n) puede estar entre 2 y $100.000 \cdot z$, en donde z es una medida de tendencia central (p. ej., media, mediana, moda) de un número esperado de moléculas duplicadas que tienen las mismas posiciones de inicio y fin. El número de combinaciones diferentes de códigos de barras moleculares puede ser al menos cualquiera de $2 \cdot z$, $3 \cdot z$, $4 \cdot z$, $5 \cdot z$, $6 \cdot z$, $7 \cdot z$, $8 \cdot z$, $9 \cdot z$, $10 \cdot z$, $11 \cdot z$, $12 \cdot z$, $13 \cdot z$, $14 \cdot z$, $15 \cdot z$, $16 \cdot z$, $17 \cdot z$, $18 \cdot z$, $19 \cdot z$, $20 \cdot z$ o $100 \cdot z$ (p. ej., límite inferior). Opcionalmente, n no es superior a $100.000 \cdot z$, $10.000 \cdot z$, $2000 \cdot z$, $1000 \cdot z$, $500 \cdot z$ o $100 \cdot z$ (p. ej., límite superior). Por lo tanto, n puede variar entre cualquier combinación de estos límites inferior y superior. El número de combinaciones puede estar entre $100 \cdot z$ y $1000 \cdot z$, $5 \cdot z$ y $15 \cdot z$, entre $8 \cdot z$ y $12 \cdot z$, o aproximadamente $10 \cdot z$. Por ejemplo, un equivalente de genoma humano haploide tiene aproximadamente 3 picogramos de ADN. Una muestra de aproximadamente 1 microgramo de ADN contiene aproximadamente 300.000 equivalentes de genoma humano haploide. El número n puede estar entre 15 y 45, entre 24 y 36, entre 64 y 2500, entre 625 y 31.000, o aproximadamente 900 y 4.000. Por ejemplo, una muestra que comprenda aproximadamente 10.000 equivalentes de genoma humano haploide de ADNfl puede codificarse con aproximadamente 36 combinaciones de seis códigos de barras moleculares diferentes. Las muestras con este tipo de código de barras pueden estar en un intervalo de aproximadamente 10 ng a cualquiera de aproximadamente 100 ng, aproximadamente 1 μ g, aproximadamente 10 μ g de polinucleótidos fragmentados, p. ej., ADN genómico, p. ej., ADNfl.

- 35 Los adaptadores son ácidos nucleicos relativamente cortos para la unión a los extremos de las moléculas de muestra para facilitar la inmovilización, amplificación, secuenciación y rastreo de ácidos nucleicos diana. Los adaptadores pueden ser bicatenarios con la posible excepción de los extremos cohesivos para facilitar la ligadura. Los adaptadores también pueden ser monocatenarios. Los adaptadores también pueden tener una porción bicatenaria y otra monocatenaria, como en los adaptadores en forma de Y y equivalentes topológicos. La longitud total de cada adaptador (medida por la cadena más larga si hay más de una) es, p. ej., menos de 250, 150, 100, 75 o 50 nucleótidos de longitud. Los adaptadores pueden incluir sitios de unión del cebador para permitir la unión de los cebadores de amplificación, extensión y/o secuenciación. Los adaptadores también pueden incluir sitios de unión de sondas de captura, tales como un oligonucleótido unido a un soporte de celda de flujo o a una perla.

- 45 Algunos adaptadores tienen una o más porciones bicatenarias y una o más porciones monocatenarias. Los adaptadores en forma de Y (véase, p. ej., el documento US 7.741.463), tallo-bucle (véase, p. ej., el documento US 10.155.939) y los adaptadores de burbujas (véase el documento US20180030532A1) son ejemplos de este tipo de adaptadores. Los adaptadores en forma de Y son ácidos nucleicos formados por dos cadenas, que están emparejadas en una porción bicatenaria (con la posible excepción de un saliente monocatenario para facilitar la ligadura), y también no emparejadas en porciones monocatenarias. Las dos porciones monocatenarias pueden representarse con la forma de la letra V unida a la porción bicatenaria, formando juntas una Y. Los adaptadores en forma de Y tienen un extremo libre en la porción bicatenaria, que puede ser un extremo romo o un extremo en donde una cadena sobresale de la otra, p. ej., por un solo nucleótido. Cada una de las cadenas simples no emparejadas tiene un extremo monocatenario. La longitud total de cada cadena de adaptadores en forma de Y es, p. ej., menos de 250, 150, 100, 75 o 50 nucleótidos de longitud. Un adaptador estándar en forma de Y de Illumina sin códigos de barras de muestra o moleculares tiene una longitud de cadena de aproximadamente 115 nucleótidos. El extremo libre de la porción bicatenaria sirve para la unión de una molécula de muestra (p. ej., mediante ligadura de extremos romos o cohesivos).

- 60 Los adaptadores de tallo-bucle (p. ej., NebNext de New England Biolabs) son similares a los adaptadores en forma de Y, salvo que las porciones monocatenarias se unen a través de un residuo de uracilo, formando así un bucle en lugar de una V. Así pues, los adaptadores de tallo-bucle son una sola cadena con un tallo formando dúplex que corresponde a la porción bicatenaria de los adaptadores en forma de Y, y un bucle que incluye dos porciones bicatenarias de ADN separadas por un uracilo (U) o una desoxiuridina (dU), que corresponden a las porciones monocatenarias de los adaptadores en forma de Y. Los residuos inmediatamente adyacentes al U o a la dU son los residuos del extremo monocatenario de las porciones monocatenarias de los adaptadores de tallo-bucle. El tallo tiene un extremo libre que puede ser romo o con cola, como en el tallo de los adaptadores en forma de Y, y se utiliza para unirse a una molécula

de muestra. Tras la unión de los adaptadores de tallo-bucle a una molécula de muestra, el U o la dU pueden eliminarse enzimáticamente dejando la misma topografía que para los adaptadores en forma de Y. La enzima USER de NEB es una mezcla de uracilo ADN glicosilasa (UDG) y la ADN glicosilasa-asa endonucleasa VIII (DGLE). La UDG cataliza la escisión de una base uracilo o desoxiuridina, formando un sitio abásico (apirimidínico) y dejando intacto el esqueleto fosfodiéster y la DGLE elimina el nucleótido abásico.

Los adaptadores de burbuja (BGI) son similares a los adaptadores de tallo-bucle y a los adaptadores en forma de Y, salvo que la región en V del adaptador en forma de Y o el bucle de los adaptadores de tallo-bucle se sustituyen por una burbuja de dos porciones monocatenarias que no forman dúplex flanqueadas a ambos lados por porciones bicatenarias. Los adaptadores de burbuja normalmente tienen dos cadenas de longitud desigual, estando parte o la totalidad de la diferencia de longitud en las porciones monocatenarias. El extremo 5' del ácido nucleico más largo tiene un nucleótido fosforilado. El extremo 3' del ácido nucleico más corto tiene normalmente un saliente desde el extremo de una porción que si no sería bicatenaria. La porción bicatenaria que contiene el nucleótido 5' fosforilado y el saliente, si está presente, se corresponde con el tallo de los adaptadores de tallo-bucle o con la porción bicatenaria de los adaptadores en forma de Y, y liga con una molécula de ácido nucleico de la muestra. Esta porción bicatenaria puede denominarse porción bicatenaria cadena abajo (en dirección 3') porque proporciona el sitio de ligadura para una molécula de muestra. La otra porción bicatenaria puede denominarse porción bicatenaria cadena arriba (en dirección 5') porque está más alejada de la molécula de muestra. Las dos cadenas sencillas en el centro que forman una burbuja se corresponden con las porciones monocatenarias que forman una V en los adaptadores en forma de Y o las porciones monocatenarias separadas por un uracilo o una desoxiuridina en los adaptadores de tallo-bucle. Los adaptadores de burbuja pueden incluir un U o una dU en la cadena más corta, cadena más larga o ambas para separar las porciones monocatenarias de la porción bicatenaria cadena arriba (en dirección 5'). Por lo general, dicho U o dU se incluye en la cadena más larga. El U o la dU pueden escindirse como en el caso de los adaptadores de tallo-bucle tras la ligadura de los adaptadores a las moléculas de muestra, dejando los adaptadores en una forma de Y.

Los sitios de unión del cebador se encuentran normalmente en las porciones monocatenarias de un adaptador en forma de Y, de tallo-bucle o de burbuja. La asimetría de las porciones monocatenarias no emparejadas permite la secuenciación específica de la cadena a partir de dos cebadores que se unen a las respectivas cadenas simples. Los adaptadores también pueden incluir sitios de unión de sondas de captura, tal como un oligonucleótido unido a un soporte de celda de flujo.

A los extremos respectivos de una molécula de ácido nucleico pueden unirse adaptadores iguales (con la posible excepción de los códigos de barras) o diferentes. Las secuencias de los adaptadores y, en particular, los segmentos para fijar la unión del cebador a una celda de flujo pueden variar dependiendo de la plataforma de secuenciación empleada.

III. Preparación de ácidos nucleicos diana

Los métodos proporcionan lecturas de secuencia directa e inversa de al menos un ácido nucleico diana y a menudo de poblaciones de al menos 100, 1000, 10.000, 100.000 o 1 M de ácidos nucleicos diana diferentes. Los ácidos nucleicos diana pueden obtenerse como parte o la totalidad de una población de ácidos nucleicos en una muestra de un sujeto o de una pluralidad de muestras de una pluralidad de sujetos o del mismo sujeto en diferentes momentos o de diferentes fuentes (es decir, tejidos o líquidos) del mismo sujeto.

Una muestra puede ser cualquier muestra biológica aislada de un sujeto. Las muestras pueden incluir tejidos corporales, tales como tumores sólidos conocidos o sospecha de estos, sangre entera, plaquetas, suero, plasma, heces, eritrocitos, glóbulos blancos o leucocitos, células endoteliales, biopsias de tejidos, líquido cefalorraquídeo, líquido sinovial, líquido linfático, líquido ascítico, líquido intersticial o extracelular, el líquido en los espacios entre las células, incluyendo líquido crevicular gingival, médula ósea, derrames pleurales, líquido cefalorraquídeo, saliva, mucosa, esputo, semen, sudor, orina. Las muestras son preferentemente líquidos corporales, en particular la sangre y sus fracciones, y la orina. Una muestra puede estar en la forma originalmente aislada de un sujeto o puede haber sido sometida a un procesamiento posterior para eliminar o añadir componentes, tales como las células, o enriquecer para un componente respecto a otro. Por lo tanto, un líquido corporal preferido para el análisis es el plasma o el suero que contiene ácidos nucleicos libres de células.

El uso de códigos de barras de muestras permite agrupar las muestras con la posterior deconvolución a partir de los códigos de barras. El número de muestras diferentes puede ser mayor o igual a 2, 5, 10, 50, 100, 500, 1000, 2000, 5000 o 10.000. El volumen de plasma puede depender de la profundidad de lectura deseada para las regiones secuenciadas. Los volúmenes de ejemplo son 0,4-40 ml, 5-20 ml, 10-20 ml. Como ejemplos, el volumen puede ser de 0,5 ml, 1 ml, 5 ml, 10 ml, 20 ml, 30 ml o 40 ml. El volumen de plasma muestreado puede ser, por ejemplo, de 5 a 20 ml.

Una muestra puede comprender diversas cantidades de ácido nucleico que contienen equivalentes genómicos. Por ejemplo, una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10.000 equivalentes de genoma humano haploide y, en el caso del ADN libre de células, aproximadamente 200.000 millones de moléculas individuales de ácido nucleico. De forma similar, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30.000 equivalentes de genoma humano haploide y, en el caso del ADN libre de células,

aproximadamente 600.000 millones de moléculas individuales. Algunas muestras contienen 1-500, 2-100, 5-150 ng de ADN libre de células, p. ej., 5-30 ng o 10-150 ng de ADN libre de células.

El ADNfl presenta un pico de fragmentos a aproximadamente 160 nucleótidos, y la mayoría de los fragmentos de este pico varían de aproximadamente 140 nucleótidos a 180 nucleótidos. En consecuencia, el ADNfl de un genoma de aproximadamente 3.000 millones de bases (p. ej., el genoma humano) puede estar compuesto por casi 20 millones (2×10^7) de fragmentos polinucleotídicos. Una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10.000 equivalentes de genoma humano haploide. (Del mismo modo, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30.000 equivalentes de genoma humano haploide). Una muestra que contenga aproximadamente 10.000 (10^4) equivalentes de genoma haploide de dicho ADN puede tener aproximadamente 200.000 millones (2×10^{11}) de moléculas de polinucleótidos individuales. Se ha determinado empíricamente que en una muestra de aproximadamente 10.000 equivalentes de genoma haploide de ADN humano, hay aproximadamente 3 polinucleótidos duplicados comenzando en cualquier posición dada. Por lo tanto, dicha colección puede contener una diversidad de aproximadamente 6×10^{10} - 8×10^{10} (aproximadamente 60.000-80.000 millones, p. ej., aproximadamente 70.000 millones (7×10^{10}) de moléculas de polinucleótidos secuenciadas de forma diferente.

Una muestra puede comprender ácidos nucleicos de diferentes tipos y orígenes. Una muestra puede contener ADN, ARN o ambos. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios o ser parcialmente bicatenarios y parcialmente monocatenarios. Una muestra puede comprender ADN de la línea germinal, ADN somático o ambos. Los ácidos nucleicos de una muestra pueden ser portadores de variaciones genéticas, los cuales pueden ser portadores de mutaciones germinales y/o somáticas. Algunas de estas mutaciones pueden ser marcadores de cáncer (p. ej., mutaciones somáticas asociadas al cáncer).

Las cantidades de ejemplo de ácidos nucleicos libres de células en una muestra antes de la amplificación varía de aproximadamente 1 fg a aproximadamente 1 µg, p. ej., de 1 pg a 200 ng, de 1 ng a 100 ng, de 10 ng a 1000 ng. Por ejemplo, la cantidad puede ser de hasta aproximadamente 600 ng, hasta aproximadamente 500 ng, hasta aproximadamente 400 ng, hasta aproximadamente 300 ng, hasta aproximadamente 200 ng, hasta aproximadamente 100 ng, hasta aproximadamente 50 ng, o hasta aproximadamente 20 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de al menos 1 fg, al menos 10 fg, al menos 100 fg, al menos 1 pg, al menos 10 pg, al menos 100 pg, al menos 1 ng, al menos 10 ng, al menos 100 ng, al menos 150 ng, o al menos 200 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de hasta 1 femtogramo (fg), 10 fg, 100 fg, 1 picogramo (pg), 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 150 ng o 200 ng de moléculas de ácido nucleico libres de células. El método puede comprender obtener de 1 femtogramo (fg) a 200 ng.

Una muestra de ejemplo es de 5-10 ml de sangre completa, plasma o suero, que incluye aproximadamente 30 ng de ADN o aproximadamente 10.000 equivalentes de genoma haploide.

Algunas muestras contienen ácidos nucleicos libres de células. Los ácidos nucleicos libres de células son ácidos nucleicos que no están contenidos dentro de una célula ni unidos a ella de otro modo o, en otras palabras, ácidos nucleicos que permanecen en una muestra de la que se extraen células intactas. Los ácidos nucleicos libres de células incluyen ADN, ARN, y sus híbridos, incluido ADN genómico, ADN mitocondrial, ARNip, miARN, ARN circulante (ARNc), ARNt, ARNr, ARN nucleolar pequeño (ARNnlp), ARN que interactúa con Piwi (ARNpi), ARN no codificante largo (ARNnc largo), o fragmentos de cualquiera de ellos. Los ácidos nucleicos libres de células pueden ser bicatenarios, monocatenarios o un híbrido de los mismos. Las moléculas de ADN bicatenario, al menos algunas de las cuales tienen salientes monocatenarios, son una forma preferida de ADN libre de células para cualquier método divulgado en el presente documento. Un ácido nucleico libre de células puede liberarse al líquido corporal a través de procesos de secreción o muerte celular, p. ej., necrosis celular y apoptosis. Algunos ácidos nucleicos libres de células se liberan en líquidos corporales a partir de células cancerosas, p. ej., ADN tumoral circulante, (ADNtc). Otros son liberados por células sanas.

Un ácido nucleico libre de células puede tener una o más modificaciones epigenéticas, por ejemplo, un ácido nucleico libre de células puede estar acetilado, metilado, ubiquitinilado, fosforilado, sumoilado, ribosilado y/o citrulinado.

Los ácidos nucleicos libres de células tienen una distribución de tamaños de aproximadamente 100-500 nucleótidos, particularmente de 110 a aproximadamente 230 nucleótidos, con un modo de aproximadamente 168 nucleótidos y un segundo pico menor en un intervalo entre 240 y 440 nucleótidos.

Los ácidos nucleicos libres de células pueden aislarse de líquidos corporales mediante una etapa de separación en la cual los ácidos nucleicos libres de células, tal y como se encuentran en solución, se separan de las células intactas y de otros componentes no solubles del líquido corporal. La separación puede incluir técnicas tales como la centrifugación o la filtración. Como alternativa, las células de los líquidos corporales pueden ser lisadas y los ácidos nucleicos celulares y libres de células procesados conjuntamente. Generalmente, después de añadir tampones y las etapas de lavado, los ácidos nucleicos pueden precipitarse con un alcohol. Pueden utilizarse etapas de limpieza adicionales, tales como columnas de sílice para eliminar contaminantes o sales. Se pueden añadir ácidos nucleicos portadores en masa no específicos, por ejemplo, puede añadirse a lo largo de la reacción para optimizar ciertos

aspectos del procedimiento, tal como el rendimiento.

Tras dicho tratamiento, las muestras pueden incluir varias formas de ácido nucleico, incluido ADN bicatenario, ADN monocatenario y ARN monocatenario. Opcionalmente, el ADN y el ARN monocatenarios pueden convertirse en formas bicatenario para que se incluyan en las etapas posteriores de procesamiento y análisis.

Los ácidos nucleicos presentes en una muestra con o sin procesamiento previo como se ha descrito anteriormente contienen normalmente una porción sustancial de moléculas en forma de moléculas parcialmente bicatenarias con salientes monocatenarios. Dichas moléculas pueden convertirse en moléculas bicatenarias con extremos romos tratándolas con una o más enzimas para proporcionar una polimerasa 5'-3' y una exonucleasa 3'-5' (o función de lectura de prueba), en presencia de los cuatro tipos de nucleótidos estándar. Dicha combinación de actividades puede extender cadenas con un extremo 3' rebajado de modo que terminen a ras con el extremo 5' de la cadena opuesta (en otras palabras, generando un extremo romo) o puede digerir cadenas con salientes 3' de modo que queden igualmente a ras con el extremo 5' de la cadena opuesta. Opcionalmente, ambas actividades pueden ser conferidas por una única polimerasa. La polimerasa es preferentemente sensible al calor, de modo que su actividad pueda interrumpirse al aumentar la temperatura. El fragmento grande de Klenow y la polimerasa T4 son ejemplos de polimerasas adecuadas.

Los ácidos nucleicos con extremos romos resultantes pueden ligarse a adaptadores con un extremo libre romo bicatenario o pueden someterse a un proceso de adición de cola para generar extremos cohesivos, que se emparejan con los correspondientes salientes monocatenarios en un extremo libre bicatenario de los adaptadores. La cola de los extremos romos puede deberse a una polimerasa que carezca de función de lectura de prueba. Esta polimerasa es preferentemente termoestable de modo que permanece activa a la temperatura elevada que desnaturaliza el uso de la polimerasa para la creación de extremos romos. Las polimerasas Taq, el fragmento grande de Bst y Tth son ejemplos de este tipo de polimerasas. La segunda polimerasa realiza una adición sin molde de un único nucleótido a los extremos 3' de los ácidos nucleicos con extremos romos. Aunque la mezcla de reacción contiene generalmente cantidades molares iguales de cada uno de los cuatro tipos de nucleótidos estándar de la etapa anterior, los cuatro tipos de nucleótidos no se añaden a los extremos 3' en proporciones iguales. Más bien, A se añade con más frecuencia, seguido de G, seguido de C y T. Dichas moléculas con cola pueden ligarse a adaptadores con una T o C complementaria por encima del extremo libre de la porción bicatenaria.

Para muestras que contengan inicialmente ácidos nucleicos bicatenarios, los presentes métodos dan lugar preferentemente a que al menos el 75, 80, 85, 90 o 95 % de los ácidos nucleicos bicatenarios de la muestra se unan a adaptadores. El ácido nucleico resultante se desnaturaliza a forma monocatenaria para realizar los presentes métodos. Preferentemente, los presentes métodos permiten secuenciar al menos el 75, 80, 85, 90 o 95 % de las moléculas inicialmente bicatenarias de la muestra.

IV. Amplificación

Los ácidos nucleicos de la muestra flanqueados por adaptadores pueden amplificarse mediante PCR y otros métodos de amplificación normalmente iniciados por cebadores que se unen a sitios de unión del cebador en adaptadores que flanquean un ácido nucleico a amplificar. Los métodos de amplificación pueden implicar ciclos de extensión, desnaturalización e hibridación resultante del termociclado o pueden ser isotérmicos como en la amplificación mediada por transcripción. Otros métodos de amplificación incluyen la reacción en cadena de la ligasa, la amplificación por desplazamiento de cadena, la amplificación basada en secuencias de ácidos nucleicos y la replicación autosostenida basada en la secuencia. La amplificación puede realizarse una o varias veces. La amplificación puede realizarse antes y de forma distinta a la secuenciación, o integrada con ésta, o ambas. La amplificación también puede realizarse antes o después del enriquecimiento de las moléculas seleccionadas de la muestra, o ambas cosas.

La amplificación en emulsión es un método comúnmente empleado para la amplificación de ácidos nucleicos en múltiples plataformas de secuenciación de nueva generación (véase, p. ej., Kanagal-Shamanna *Methods Mol Biol.* 2016;1392:33-42. doi: 10.1007/978-1-4939-3360-0_4. PMID: 26843044). El principio básico de la amplificación en emulsión es la dilución y compartimentación de los ácidos nucleicos a amplificar en gotitas de agua en una emulsión de agua en aceite. Preferentemente, la dilución se realiza hasta un grado en donde cada gotita contiene un único ácido nucleico diana y funciona como un microrreactor. La compartimentación de los ácidos nucleicos individuales en distintas gotitas de reacción permite su amplificación independientemente unos de otros, reduciendo de este modo la formación de subproductos quiméricos y el sesgo general de amplificación. Los ácidos nucleicos sometidos a amplificación pueden ser el resultado de fragmentos previos de moléculas mayores, tal como, mediante la digestión de una muestra de ADN genómico. Los ácidos nucleicos pueden unirse a adaptadores antes de la amplificación y la amplificación puede proporcionar sitios para la unión de cebadores. Los ácidos nucleicos, normalmente flanqueados por adaptadores, están inmovilizados en soportes, antes y/o durante la amplificación. Los soportes son habitualmente perlas sustancialmente esféricas u otras partículas, dimensionados para que una sola gotita pueda alojar un soporte. La amplificación requiere cebador(es), sales, una polimerasa y nucleótidos trifosfato como en otras formas de amplificación. La amplificación se configura de tal manera que una población clonal de amplicones de la misma molécula de ácido nucleico original se adhiere al mismo soporte dentro de un compartimento. Esta configuración puede conseguirse uniendo varias copias de un cebador directo a cada soporte. El cebador directo es complementario de un adaptador en el extremo 5' de una cadena de ácido nucleico a amplificar. Esta cadena y las copias de la misma

generadas durante la amplificación pueden hibridarse con el cebador directo del soporte. La extensión del cebador genera una cadena complementaria unida covalentemente al cebador, que a su vez está unido al soporte. El resultado final es una población clonal de ácido nucleico monocatenario unido mediante un adaptador 5' a un soporte. Las diferentes poblaciones clonales se unen a diferentes soportes. Tras la amplificación, la emulsión puede romperse mezclando los soportes. En la población resultante de soportes mezclados pueden realizarse otras reacciones de extensión y/o secuenciación, o bien los soportes pueden redistribuirse primero a pocillos individuales o similares de celdas de flujo de secuenciación.

V. Enriquecimiento

Las poblaciones de ácidos nucleicos de muestra pueden someterse a enriquecimiento para los ácidos nucleicos diana de interés. El enriquecimiento puede realizarse mediante purificación por afinidad, p. ej., mediante hibridación con oligonucleótidos inmovilizados complementarios de las secuencias de interés. El enriquecimiento puede realizarse antes o después de la ligadura a los adaptadores, y antes o después de la amplificación, o cualquier combinación de los mismos.

VI. Secuenciación

El término "secuenciación" se refiere a un proceso para generar una secuencia de un ácido nucleico diana. La secuenciación puede realizarse en moléculas individuales (secuenciación de molécula única) o en poblaciones clonales originadas a partir de la misma molécula (secuenciación clonal). La secuenciación puede formar parte de una extensión de cebador dirigida por molde (secuenciación por síntesis) en donde se detectan los nucleótidos sucesivos incorporados a una cadena naciente. La secuenciación también puede ser independiente de la extensión, tal como en la secuenciación por hibridación o ligadura. La secuencia generada a partir de un único molde o población clonal de los mismos moldes puede denominarse lectura de secuenciación. Una lectura de secuenciación puede incluir la secuencia de parte o la totalidad de un ácido nucleico diana, así como algunos o todos los adaptadores que flanquean al ácido nucleico diana. Por ejemplo, una lectura de secuenciación puede incluir al menos 5, 10, 15, 20, 50, 100, 250 o 500 nucleótidos de secuencia de un ácido nucleico diana. Los nucleótidos de una lectura de secuencia pueden corresponder a nucleótidos contiguos en un ácido nucleico diana o pueden tener tramos contiguos separados por hueco(s) de una o más posiciones en las que no se ha determinado la identidad de un nucleótido. Es preferible que los huecos no ocupen más del 1 %, 5 % o 10 % de las posiciones de una lectura de secuenciación. Una lectura de secuenciación puede formarse a partir de los tipos de nucleótidos presentes en un ácido nucleico diana monocatenario o, como alternativa, formarse a partir de sus tipos de nucleótidos complementarios. Una lectura de secuenciación directa e inversa se refiere a la secuencia generada en direcciones opuestas con respecto a un ácido nucleico diana monocatenario. La lectura de secuenciación directa se determina 3'-5' con respecto a un molde monocatenario y la lectura de secuenciación inversa 3'-5' con respecto al complemento del molde o 5'-3 con respecto al propio molde.

La pirosecuenciación 454 es un ejemplo de un método de secuenciación (Siqueira *et al.*, J Oral Microbiol. 2012;4:10.3402/jom.v4i0.10743. doi:10.3402/jom.v4i0.10743). Los ácidos nucleicos individuales que se van a secuenciar se ligan a adaptadores y se anclan a perlas individuales, preferentemente un fragmento original por perla. A continuación, los ácidos nucleicos se amplifican en las perlas de forma que se forma una población clonal en las perlas. La amplificación va seguida de la detección química de las reacciones de síntesis de ADN iniciadas por los amplicones en una cámara en miniatura en donde se mide la liberación de pirofosfato. Inundando consecutivamente dicha cámara con reactivos de secuenciación que contengan uno de los 4 nucleótidos estándar, cuando se incorpora el nucleótido correcto en la cadena sintetizada, se mide la liberación de pirofosfato utilizando una reacción generadora de luz. La intensidad de la luz también proporciona información sobre las series homopoliméricas de nucleótidos en la secuencia. La pirosecuenciación fue desarrollada por Pyrosequencing AB y posteriormente adquirida por Qiagen, que la licenció a 454 Life Sciences, antes de ser finalmente adquirida por Roche.

Ion Torrent es otro ejemplo de método de secuenciación (véase Hu *et al.*, Human Immunology 82, 801-811 (2021)). Los ácidos nucleicos que se van a secuenciar se unen a adaptadores. A continuación, los fragmentos de ADN adaptados se anclan a perlas, preferentemente una perla para un fragmento. A continuación, los fragmentos se amplifican en las perlas mediante amplificación en emulsión para generar grandes poblaciones de perlas, cada una de ellas con una población clonal de la misma molécula original. A continuación, las perlas fluyen a través del chip que contiene los pocillos configurados de forma que solo pueda entrar una perla en cada pocillo. A continuación, los reactivos de secuenciación fluyen por los pocillos, cuando se incorpora el nucleótido apropiado, se emite un ion hidrógeno y se registra la señal. La incorporación de nucleótidos se convierte directamente en voltaje que se registra directamente. El sistema Ion Torrent es comercializado por Thermo-Fisher y existen varias versiones de la plataforma, incluido el sistema Ion Personal Genome Machine™ (PGM™), Ion Proton™ System, Ion S5 system e ION S5 XL system, cada uno con características de rendimiento diferentes (véase la página web thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing.html). También está disponible un sistema automatizado de preparación de bibliotecas y moldes (Ion Chef™). Admite un gran número de aplicaciones, incluida la secuenciación dirigida y de novo de ADN y ARN, secuenciación del transcriptoma, secuenciación microbiana, detección de variaciones en el número de copias, secuenciación de ARN pequeño y miARN y CHIP-seq (secuenciación por inmunoprecipitación de la cromatina).

La secuenciación de Illumina se basa en una técnica conocida como amplificación en puente en la cual los ácidos nucleicos con adaptadores apropiados ligados en cada extremo se utilizan como sustratos para reacciones de síntesis de amplificación repetidas en un soporte sólido (portaobjetos de vidrio) que contiene cebadores complementarios de un adaptador ligado (véase, p. ej., Slatko *et al.*, Curr. Protoc. Mol. Biol. 122(1), e59 (2018)). Los cebadores en el portaobjetos están espaciados de tal manera que los ácidos nucleicos, que luego se someten a repetidas rondas de amplificación, crean "grupos" clonales formados por aproximadamente 1000 copias. Cada portaobjetos de vidrio puede soportar millones de reacciones de grupo en paralelo. Durante las reacciones de síntesis, se incorporan nucleótidos modificados patentados, correspondientes a cada una de las cuatro bases, cada uno con un marcador fluorescente diferente, y a continuación se detectan. Los nucleótidos también actúan como terminadores de la síntesis de cada reacción, que se desbloquean tras su detección para la siguiente ronda de síntesis.

Ultima Genomics ha desarrollado una tecnología de secuenciación por síntesis para colonias clonales de un ácido nucleico diana situadas en una plataforma giratoria. Las moléculas de ácido nucleico diana se hibridan con un cebador, que puede ampliarse con una mezcla de nucleótidos estándar y de terminación. La rotación de la plataforma puede configurarse para coordinar los ciclos de exploración con los ciclos de extensión (véase, p. ej., los documentos US20190153520, US20190153531).

Otros métodos de secuenciación son, por ejemplo, secuenciación de Sanger, secuenciación en tiempo real de una sola molécula (Pac-Bio), secuenciación con ONT, secuenciación de exones, secuenciación basada en microscopía electrónica, secuenciación de paneles, secuenciación mediada por transistor, secuenciación directa, secuenciación aleatoria de escopeta, secuenciación de todo el genoma, electroforesis capilar, electroforesis en gel, secuenciación dúplex, secuenciación por ciclos, co-amplificación a baja temperatura de desnaturalización-PCT (COLD-PCR), secuenciación por terminador de colorante reversible, secuenciación de extremos emparejados, secuenciación a corto plazo, secuenciación por exonucleasa, secuenciación por ligamiento, secuenciación de lectura corta, secuenciación por nanoporos, secuenciación de semiconductores, secuenciación por ligadura, secuenciación por hibridación, ARN-Seq (Illumina), Digital Gene Expression (Helicos), secuenciación de nueva generación, secuenciación de moléculas individuales por síntesis (SMSS) (Helicos), secuenciación masiva paralela, secuenciación de escopeta, Oxford Nanopore, Roche Genia, secuenciación de Maxam-Gibert, paseo con cebador, SOLiD, secuenciación MS-PET o plataformas Nanopore, y combinaciones de las mismas. Las reacciones de secuenciación pueden realizarse en diversas unidades de procesamiento de muestras, que pueden tener múltiples carriles, múltiples canales, múltiples pocillos, u otro medio para procesar múltiples conjuntos de muestras de forma sustancialmente simultánea. La unidad de procesamiento de muestras también puede incluir múltiples cámaras de muestra para permitir el procesamiento de múltiples análisis simultáneamente.

Las reacciones de secuenciación pueden proporcionar una cobertura de secuencia del genoma de al menos el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 %. En otros casos, la cobertura de la secuencia del genoma puede ser inferior al 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 %.

Las reacciones de secuenciación simultánea pueden realizarse mediante secuenciación multiplex. En algunos casos, los amplicones de los ácidos nucleicos de muestra pueden secuenciarse con al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10.000, 50.000, 100.000 reacciones de secuenciación. En otros casos, los amplicones de los ácidos nucleicos de la muestra pueden secuenciarse con menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10.000, 50.000, 100.000 reacciones de secuenciación. Las reacciones de secuenciación pueden realizarse secuencial o simultáneamente. El análisis posterior de los datos puede realizarse sobre la totalidad o parte de las reacciones de secuenciación. En algunos casos, el análisis de datos puede realizarse en al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10.000, 50.000, 100.000 reacciones de secuenciación. En otros casos, el análisis de datos puede realizarse en menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10.000, 50.000, 100.000 reacciones de secuenciación.

El método de secuenciación puede ser la secuenciación paralela masiva, es decir, secuenciación simultánea (o en rápida sucesión) de cualquiera de al menos 100, 1000, 10.000, 100.000, 1 millón, 10 millones, 100 millones o 1.000 millones de ácidos nucleicos diana diferentes.

Las muestras pueden dividirse en dos o más alícuotas antes o después de agruparlas para el análisis de la modificación del ADN (véase, p. ej., Gouil *et al.*, Essays Biochem. 63(6):639-648 (2019)). Una alícuota de las muestras se trata de forma que los nucleótidos no modificados sufran una sustitución por un nucleótido diferente. Por ejemplo, en la secuenciación por bisulfito sódico, las citosinas no modificadas pueden convertirse en uracilo, mientras que las citosinas metiladas no se modifican. La comparación de las lecturas de secuenciación de las distintas alícuotas indica qué citosinas fueron objeto de modificación.

VII. Análisis y deconvolución de códigos de barras

Las lecturas de secuenciación pueden segregarse de acuerdo con su muestra de origen mediante deconvolución basada en los códigos de barras de las muestras. Las lecturas de secuenciación pueden segregarse en familias que representan copias de amplificación de la misma molécula original a partir de los códigos de barras moleculares,

normalmente a partir de una combinación de códigos de barras moleculares cadena arriba (en dirección 5') y cadena abajo (en dirección 3'), y a veces de la secuencia del ácido nucleico de la muestra. Si se utiliza un código de barras molecular único, el código de barras molecular o la combinación de códigos de barras moleculares cadena arriba (en dirección 5') y cadena abajo (en dirección 3') es suficiente para indicar la familia de origen (es decir, todas las lecturas de secuenciación tienen la misma combinación de códigos de barras), opcionalmente tras permitir la corrección de errores de secuenciación en el código de barras, incluidos los complementos de la cadena opuesta se agrupan en la misma familia). Si se utiliza un código de barras no único, entonces las familias se identifican basándose en que tienen los códigos de barras moleculares o la misma combinación de códigos de barras moleculares junto con una propiedad de igualdad entre las secuencias de moléculas de muestra (tal como los mismos puntos de inicio y parada, o la misma longitud) cuando se alinean con una secuencia de referencia conocida. Las lecturas de secuenciación dentro de la misma familia pueden incluir lecturas de secuenciación de una o ambas cadenas de la misma molécula original.

Las lecturas de secuenciación de los miembros de la familia pueden compilarse para obtener nucleótido(s) de consenso en posiciones especificadas o secuencia de consenso en algunas o todas las posiciones de una molécula de ácido nucleico en la muestra original. Si los miembros de una familia incluyen lecturas de secuenciación de cadenas opuestas, las secuencias de una cadena pueden convertirse en sus complementos con el fin de compilar y alinear todas las lecturas de secuenciación para obtener nucleótido(s) o secuencias de consenso. Un tipo de nucleótido de consenso en una posición puede definirse como el tipo de nucleótido que más frecuentemente ocupa esa posición entre las lecturas de secuenciación alineadas. Del mismo modo, una secuencia de consenso puede definirse como la secuencia de dichos tipos de nucleótidos de consenso. Para que un tipo de nucleótido se denomine de consenso en una posición determinada en las lecturas de secuenciación alineadas, también se puede requerir que el tipo de nucleótido aparezca por encima de un nivel de frecuencia umbral entre los tipos de nucleótidos que ocupan esa posición en las lecturas de secuenciación alineadas. Por ejemplo, puede requerirse que el tipo de nucleótido esté presente en esa posición en al menos el 50, 60, 70, 80 o 90 % de las lecturas de secuenciación. Puede requerirse adicional o como alternativa que el tipo de nucleótido esté presente en al menos una lectura de secuenciación de ambas cadenas de una molécula original. Puede requerirse adicional o como alternativa que el tipo de nucleótido no sea contradicho por más de un número umbral de lecturas de secuenciación de una o ambas cadenas en las que la posición alineada esté ocupada por un tipo de nucleótido diferente. Opcionalmente, las posiciones que muestren discordancia entre las lecturas de secuenciación directa e inversa por pares de la misma molécula pueden descartarse en esas lecturas al compilar una secuencia de consenso. Las deleciones o inserciones de consenso pueden identificarse mediante análisis similares de representación y/o presencia en ambas cadenas o sustituciones.

Algunas familias pueden incluir una sola lectura de secuenciación. En este caso, esta secuencia puede tomarse como la secuencia de un ácido nucleico en la muestra antes de la amplificación. Como alternativa, las familias con una única secuencia miembro pueden eliminarse del análisis posterior.

Los criterios descritos anteriormente para identificar nucleótidos o secuencia de consenso ayudan a filtrar las variaciones genuinas de nucleótidos a partir de una secuencia de referencia en moléculas de muestra originales y las variaciones resultantes de errores de amplificación o secuenciación. Es probable que las variaciones de ácidos nucleicos presentes en las moléculas de muestra originales tengan una mayor representación en las lecturas de secuenciación en general y, en particular, en las lecturas de secuenciación de ambas cadenas que las variaciones resultantes de errores de amplificación o secuenciación y, por tanto, se designen como tipos de nucleótidos o secuencia de consenso de dichos tipos de nucleótidos.

Habiendo determinado los nucleótidos de consenso y/o las secuencias de consenso dentro de familias individuales, los resultados pueden compilarse para proporcionar una indicación de las variaciones de nucleótidos presentes en una muestra en comparación con una secuencia de referencia conocida. La secuencia de referencia conocida puede ser la de un gen, cromosoma o genoma, entre otros. Dicha compilación puede proporcionar un filtro adicional para distinguir las variaciones de secuencia genuinas de los errores de amplificación y secuenciación y proporcionar una indicación de la representación o frecuencia alélica de dichas variaciones en relación con el tipo silvestre en una muestra. Para cualquier posición de interés en una secuencia de referencia para una muestra (p. ej., secuencia del genoma humano de tipo silvestre), se puede determinar qué familias tienen lecturas de secuenciación que abarcan esa posición. En esas familias se puede determinar una representación del tipo de nucleótido variante, deleciones o inserciones, dado el caso, y el tipo de nucleótido silvestre para esa posición. Se puede considerar que una variación está presente en la posición si el número de familias que incluye un tipo de nucleótido variante, deleciones o inserciones supera un umbral, o la proporción de familias con el tipo de nucleótido variante, deleción o inserción respecto al tipo silvestre supera un umbral entre otros criterios. La proporción entre el tipo de nucleótido variante, deleción o inserción respecto al tipo de nucleótido silvestre también proporciona una indicación de la representación del nucleótido variante. Dicho análisis puede realizarse para cada nucleótido de interés en una secuencia de referencia correspondiente a una muestra concreta, proporcionando así un perfil de variantes de esa muestra. El análisis puede repetirse para cada muestra utilizando familias de lecturas de secuenciación y sus nucleótidos o secuencias de nucleótidos de consenso obtenidas como se ha comentado anteriormente. Por lo tanto, cada muestra puede caracterizarse por un perfil de tipo de nucleótido variante.

Los nucleótidos o secuencias de consenso también pueden compararse entre diferentes alícuotas de muestras sometidas a un tratamiento que dé lugar a una sustitución diferencial de nucleótidos modificados y no modificados,

como en el análisis de bisulfito. Este análisis indica qué nucleótidos de las moléculas de muestra están modificados, tal como por metilación.

Las familias de secuencias también pueden utilizarse para proporcionar una indicación de la variación del número de copias (véase, p. ej., los documentos WO2017/106768, WO/2015/100427). El número de familias que tienen una lectura de secuenciación de consenso que abarca un locus particular o dentro de una ventana definida de un genoma en comparación con el número de familias en las que está cartografiado un locus o ventana en otra parte del genoma, proporciona una medida de la variación del número de copias, que puede surgir tanto de la amplificación como de la pérdida de un alelo. El número de familias medido puede normalizarse según sea necesario para tener en cuenta factores como las diferencias en el tamaño de la ventana, cobertura de secuenciación o enriquecimiento para diferentes regiones de un genoma.

VIII. Aplicaciones

Las secuencias y las variaciones genéticas en ellas determinadas por los presentes métodos pueden utilizarse para diagnosticar la presencia o ausencia de afecciones, particularmente cáncer, en un sujeto, para caracterizar afecciones (p. ej., selección del tratamiento adecuado o estadificación del cáncer o determinación de la heterogeneidad de un cáncer), controlar la respuesta al tratamiento de una afección, efecto pronóstico del riesgo de desarrollar una afección o evolución posterior de una afección.

Con los presentes métodos pueden detectarse diversos tipos de cáncer. Las células cancerosas, como la mayoría de las células, pueden caracterizarse por la tasa de renovación, en la cual las células viejas mueren y son sustituidas por células nuevas. Generalmente las células muertas, en contacto con la vasculatura en un sujeto determinado, pueden liberar ADN o fragmentos de ADN en el torrente sanguíneo. Lo mismo ocurre con las células cancerosas en los distintos estadios de la enfermedad. También se pueden caracterizar las células cancerosas, en función del estadio de la enfermedad, por diversas aberraciones genéticas tales como la variación del número de copias, así como por mutaciones raras. Este fenómeno puede utilizarse para detectar la presencia o ausencia de individuos con cáncer mediante los métodos descritos en el presente documento.

Entre los tipos y el número de cánceres que pueden detectarse figuran las neoplasias hemáticas, cánceres de cerebro, cánceres de pulmón, cánceres de piel, cánceres de nariz, cánceres de garganta, cánceres de hígado, cánceres óseos, linfomas, cánceres de páncreas, cánceres de piel, cánceres de intestino, cánceres rectales, cánceres de tiroides, cánceres de vejiga, cánceres de riñón, cánceres de boca, cánceres de estómago, tumores sólidos, tumores heterogéneos, tumores homogéneos y similares.

Los cánceres pueden detectarse a partir de variaciones genéticas, incluidas las mutaciones, mutaciones raras, indeles, variaciones del número de copias, transversiones, translocaciones, inversión, deleciones, aneuploidía, aneuploidía parcial, poliploidía, inestabilidad cromosómica, alteraciones de la estructura cromosómica, fusiones génicas, fusiones cromosómicas, truncamientos de genes, amplificación génica, duplicaciones de genes, lesiones cromosómicas, lesiones del ADN, cambios anormales en las modificaciones químicas de los ácidos nucleicos, cambios anormales en los patrones epigenéticos, cambios anormales en la metilación de ácidos nucleicos infección y cáncer.

Los datos genéticos también pueden utilizarse para caracterizar una forma específica de cáncer. Los cánceres son frecuentemente heterogéneos tanto en su composición como en su estadificación. Los datos del perfil genético pueden permitir la caracterización de subtipos específicos de cáncer que pueden ser importantes para el diagnóstico o el tratamiento de ese subtipo específico. Esta información también puede proporcionar a un sujeto o a un profesional pistas sobre el pronóstico de un tipo específico de cáncer y permitir a un sujeto o a un profesional adaptar las opciones de tratamiento en función de la evolución de la enfermedad. Algunos cánceres progresan, volviéndose más agresivos y genéticamente inestables. Otros cánceres pueden seguir siendo benignos, inactivos o latentes. El sistema y los métodos de esta divulgación pueden ser útiles para determinar la progresión de la enfermedad.

Los presentes métodos también son útiles para determinar la eficacia de determinadas opciones de tratamiento. Por ejemplo, el número de variaciones detectadas, independientemente de su identidad exacta, es un predictor de la susceptibilidad a la inmunoterapia porque las mutaciones crean neoepítopos que pueden ser objeto de ataque inmunitario (véase, p. ej., el documento US20200370129).

Otras variaciones o variaciones del número de copias indican la idoneidad de un fármaco concreto. Algunos ejemplos de estas variaciones son los siguientes:

Gen	Variación	Cáncer	Fármaco
EGFR/ErbB1	Mutaciones (p. ej., L858R, ex19del, T790M)	CPNM	gefitinib, erlotinib, afatinib, osimertinib, dacomitinib

ES 3 020 408 T3

(continuación)			
Gen	Variación	Cáncer	Fármaco
HER2/ErbB2	Amplificación Amplificación Mutaciones puntuales (V659E)	Mama Esofagogástrico	trastuzumab, T-DM1, trastuzumab + pertuzumab, lapatinib, neratinib
c-Met	Mutaciones de salto del exón 14, amplificación	CPNM CPNM	trastuzumab lapatinib crizotinib, capmatinib, savolitinib*, tepotinib
RET	Fusión	CPNM	selpercatinib, pralsetinib, cabozantinib, 3 A vandetanib
ALK	Fusión	CPNM	crizotinib, alectinib, ceritinib, lorlatinib brigatinib
	Mutaciones (F1196M, F1196Q)	Sarcoma de tejidos blandos	crizotinib, ceritinib
ROS1	Fusión, mutación	CPNM	crizotinib, entrectinib
NTRK	Fusión	Todos los tumores	larotrectinib, entrectinib
c-Kit	Mutaciones (p. ej., 449_514mut), deleciones (p. ej., D419del)	TEGI	imatinib, sunitinib, regorafenib, sorafenib
		Tumores tímicos	sunitinib
	Mutaciones (p. ej., K642E)	Melanoma	imatinib
PDGFR	Mutaciones (p. ej., D842V), deleciones (p. ej., C456_N468del)	TEGI	imatinib, dasatinib
		Leucemia, mielodisplasia	imatinib
FGFR1	Amplificación	CCEL	erdafitinib
		CPNM	AZD4547
FGFR2	Fusión, mutación	Vejiga, colangiocarcinoma	erdafitinib, pemigatinib
	Amplificación	Mama	dovitinib
FGFR3	Fusión, mutación	Vejiga	erdafitinib
RAS	Tipo silvestre	CCR	cetuximab, panitumumab
BRAF	Mutaciones (p. ej., V600E)	Melanoma	vemurafenib, dabrafenib, trametinib, trametinib
		CPNM	dabrafenib + trametinib
		Histiocitosis	cobimetinib
	Mutación (V600E)	CCR	encorafenib + cetuximab
	Fusiones	Ovárico	trametinib, cobimetinib
MEK	Mutaciones	Melanoma, CPNM, ovárico, trastorno histiocítico	trametinib, cobimetinib, selumetinib
mTOR	Mutaciones (p. ej., E2014K)	Vejiga, Cáncer de células renales	everolimus, temsirolimus
AKT	Mutación (E17K)	Mama, ovárico	capivasertib
PTEN	Deleciones homocigotas, mutaciones de pérdida de función	Mama	capivasertib
PIK3CA	Mutaciones	Mama	alpelisib
CDK4	Amplificación	Sarcoma de tejidos blandos	palbociclib
IDH1	Mutaciones	LMA, colangiocarcinoma	ivosidenib
IDH2	Mutaciones	LMA	enasidenib

(continuación)

Gen	Variación	Cáncer	Fármaco
BRCA1/2 y ATM	Mutaciones (somáticas)	Mama	olaparib, talazoparib, rucaparib
	Mutaciones (somáticas)	ovárico, próstata	rucaparib, olaparib
REa	Mutaciones (p. ej., E380Q)	Mama	fulvestrant
MSI-H	No aplicable	Todos	pembrolizumab
TML	No aplicable	Múltiples tipos de tumores	pembrolizumab, nivolumab

Los métodos actuales también pueden utilizarse para hacer seguimiento de la terapia. Por ejemplo, un tratamiento exitoso puede asociarse inicialmente a un aumento de las variaciones de nucleótidos o del número de copias en el ADN libre de células, a medida que las células cancerosas mueren y liberan su ADN a la circulación. Este aumento inicial puede ir seguido de una disminución que refleje que quedan menos células cancerosas, si es que quedan, que liberan su ADN. También puede producirse un aumento posterior de las variaciones de los nucleótidos o del número de copias tras un período de remisión, lo que constituye un indicio de recurrencia del cáncer.

Los presentes métodos también pueden utilizarse para detectar variaciones genéticas en afecciones distintas del cáncer. Las células inmunitarias, tales como linfocitos B, sufren variaciones en el número de copias asociadas a determinadas enfermedades. Las expansiones clonales pueden controlarse mediante la detección de variaciones en el número de copias como medida de la progresión de la enfermedad. Los presentes métodos pueden utilizarse para determinar o perfilar las actividades de rechazo del organismo hospedador, a medida que las células inmunitarias intentan destruir el tejido trasplantado para controlar el estado del tejido trasplantado, así como alterar el curso del tratamiento o la prevención del rechazo. La variación del número de copias o la variante nucleotídica pueden utilizarse para determinar cómo cambia una población de patógenos durante el curso de la infección. Por ejemplo, durante las infecciones crónicas, tales como las infecciones por VIH/SIDA o hepatitis, y los virus pueden cambiar de estado de ciclo vital y/o mutar a formas más virulentas en el curso de la infección.

Los presentes métodos pueden utilizarse para generar o perfilar, obtener la huella genética o conjunto de datos que es la suma de información genética obtenida de diferentes células en una enfermedad heterogénea. Este conjunto de datos puede comprender la variación del número de copias y la variación de nucleótidos, o ambas.

Los presentes métodos se pueden usar para diagnosticar, pronosticar, controlar u observar cánceres u otras enfermedades de origen fetal. Es decir, estas metodologías pueden emplearse en una mujer embarazada para diagnosticar, pronosticar, controlar u observar cánceres u otras enfermedades en un sujeto no nacido cuyo ADN y otros ácidos nucleicos pueden co-circular con moléculas maternas.

I. Kits

Cualquiera o todos los reactivos descritos para llevar a cabo los métodos anteriores pueden incluirse en un kit. Algunos de estos kits incluyen un cebador con una región de unión a la diana 3' y una región de tallo-bucle 5', como se ha descrito anteriormente. Estos kits también pueden incluir un cebador de bucle, uno o más adaptadores para la unión a un ácido nucleico diana y un cebador de secuenciación inversa, cada uno como se ha descrito anteriormente. Como se ha descrito anteriormente, los métodos utilizan adaptadores 5' y 3' que flanquean un ácido nucleico diana, que incluyen diferentes sitios de unión del cebador. Otros kits incluyen un cebador con dos regiones de unión a la diana 3' separadas por una región intermedia. Dichos cebadores también pueden incluir uno o más adaptadores y un cebador de secuenciación inversa como se ha descrito anteriormente.

Los adaptadores de estos kits pueden suministrarse por separado, p. ej., un adaptador de 5' y otro de 3'. Como alternativa, puede utilizarse un único adaptador (con la posible excepción de la(s) región(es) del código de barras) con regiones monocatenarias asimétricas, p. ej., forma de Y, tallo-horquilla o burbuja. Dicho adaptador, cuando flanquea ambos extremos de un ácido nucleico diana bicatenario, está dispuesto para flanquear las cadenas individuales del ácido nucleico diana con las diferentes secuencias sencillas del adaptador. Estas secuencias monocatenarias que flanquean un ácido nucleico diana monocatenario pueden considerarse adaptadores 5' y 3'.

Estos kits también pueden incluir una polimerasa, opcionalmente, una polimerasa con actividad de desplazamiento de la cadena, y nucleótidos trifosfato para la extensión. Dicho kit puede incluir una fosfatasa para eliminar un grupo bloqueador de fosfato de un cebador. Dichos nucleótidos trifosfato o al menos algunos de ellos pueden estar etiquetados y/o algunos pueden ser análogos de nucleótidos que dan lugar a la terminación de la cadena. Por lo tanto, por ejemplo, dicho kit puede incluir un conjunto de nucleótidos trifosfato estándar y un conjunto de nucleótidos trifosfato etiquetados y/o nucleótidos trifosfato terminadores de cadena. Dicho kit también puede incluir un prospecto u otras instrucciones para llevar a cabo los métodos, que puede facilitarse en formato impreso, tal como el papel, o de forma blanda para su visualización en un monitor de ordenador, o similares.

Ejemplos

Una forma ilustrativa del método se muestra en la figura 1. La figura 1 a la izquierda muestra un producto de PCR en emulsión, es decir, moléculas de ADN diana monocatenario flanqueadas por un par de adaptadores. Las moléculas de ADN monocatenario unidas a la misma perla son amplicones clónicos de la misma molécula original. Las perlas se encuentran inicialmente en pocillos separados dentro de una emulsión, pero ésta puede romperse mezclando las perlas antes de continuar. La derecha de la figura 1 muestra un tallo-bucle (o cebador de horquilla). Una porción 3' del cebador es una región de unión a la diana configurada para unirse al adaptador en el extremo 3' de las moléculas de ADN unidas a las perlas. La porción 5' del cebador es un tallo-bucle. La figura 2 muestra el cebador de tallo-bucle hibridado con el adaptador en el extremo 3' de las moléculas de ADN. La figura 3 muestra la ligadura del extremo 5' del cebador de tallo-bucle con el extremo 3' de las moléculas de ADN unidas a la perla. La ligadura puede realizarse como alternativa después de la secuenciación directa. La figura 4 muestra la extensión del extremo 3' del cebador de tallo-bucle en una reacción de secuenciación por síntesis para generar una cadena complementaria y proporcionar una lectura de secuenciación directa. La figura 5 muestra la hibridación de un cebador con la región de bucle del cebador de tallo-bucle. La figura 6 muestra la extensión desde el cebador hasta la región de bucle. La extensión desplaza la cadena complementaria sintetizada previamente por extensión del extremo 3' del cebador de tallo-bucle. La figura 7 muestra un cebador de lectura inversa que hibrida con la región adaptadora 3' de la cadena complementaria. La figura 8 muestra la extensión del cebador de lectura inversa con la cadena de complementariedad actuando como molde para generar una lectura de secuenciación de la cadena inversa.

La figura 10 muestra una disposición similar a la figura 1, pero en este caso el adaptador en el extremo 3' del ácido nucleico diana tiene dos sitios de unión del cebador separados por un marcador de muestra. El sitio de unión del cebador proximal al extremo 3' de la molécula de ADN diana se denomina sitio de unión del cebador directo, y el sitio de unión del cebador distal al extremo 3' de la molécula de ADN diana (es decir, 3' respecto al marcador de muestra en la figura 10) se denomina sitio de unión del cebador al marcador. La inclusión de un marcador de muestra entre los sitios de unión del cebador es opcional. La figura 11 muestra un cebador funcional doble que incluye dos regiones de unión a la diana configuradas para formar dúplex con el sitio de unión del cebador directo y el sitio de unión del cebador al marcador en el adaptador del ácido nucleico diana. La región de unión a la diana complementaria del sitio de unión del cebador al marcador puede bloquearse en su extremo 3', p. ej., con un grupo fosfato para evitar su extensión hasta que sea necesario. Otras moléculas bloqueantes que pueden utilizarse incluyen un grupo amonio, un nucleótido modificado, p. ej., un 3' 2' di desoxinucleótido o 3' desoxiadenosina 5'-trifosfato (cordicepina o una secuencia corta de nucleótidos con una orientación de 3' a 5', para que no haya ningún grupo hidroxilo libre en el extremo 3', el uso de un grupo alquilo 3', una fracción 3' no nucleotídica), fosforotioato, restos de alcano-diol, ácido peptidonucleico (PNA), residuos de nucleótidos que carecen de un grupo hidroxilo 3' en el extremo 3', o una proteína de unión a ácidos nucleicos (véase, p. ej., Arnold *et al.*, "Non-Nucleotide Linking Reagents for Nucleotide Probes", la patente de Estados Unidos n.º 6.031.091).

Cada una de las regiones de unión a la diana tiene un extremo 3' libre para permitir la extensión a partir del cebador. Las regiones de unión a la diana flanquean una región intermedia, que incluye una modificación por inversión, tal como un nucleótido invertido para permitir que las regiones de unión a la diana estén en orientaciones opuestas y, por tanto, ambas tengan extremos 3' libres. La parte izquierda de la figura muestra el cebador funcional doble con ambos sitios de unión a la diana en dúplex con los correspondientes sitios de unión del cebador en el adaptador. La figura 12 muestra la extensión de la región de unión a la diana formando un dúplex con el sitio de unión del cebador directo. La extensión puede ser una reacción de secuenciación por síntesis en la cual la identidad de los nucleótidos añadidos se detecta sucesivamente generando una lectura de secuenciación directa o puede ser una reacción de extensión sin secuenciación. Tras la síntesis de una cadena complementaria del ácido nucleico diana monocatenario, el bloqueo con fosfato del extremo 3' del sitio de unión a la diana del marcador puede eliminarse mediante el tratamiento con cinasa (figura 13). La figura 14 muestra la extensión del sitio de unión a la diana formando un dúplex con el sitio de unión del cebador con el marcador. Si esta región de unión a la diana se bloquea inicialmente para evitar una extensión prematura, se puede eliminar el bloqueo, p. ej., con tratamiento de fosfatasa para un bloqueo de fosfato. La extensión puede proporcionar una lectura del marcador de muestra. La extensión también genera otra cadena complementaria del ácido nucleico diana inicialmente monocatenario que desplaza a la cadena complementaria original, como se muestra en la figura 15. La extensión puede realizarse con la polimerasa con actividad de desplazamiento de cadena. Como alternativa, la extensión puede realizarse con una polimerasa que no desplace la cadena y que tenga actividad exonucleasa 5'-3' (p. ej., Taq), en cuyo caso no es necesario bloquear el terminal 3' y eliminarlo antes de la extensión. La cadena complementaria original sigue anclada a la perla mediante el cebador funcional doble y el ácido nucleico diana original. La figura 16 muestra un cebador inverso que se une al adaptador en el extremo 3' de la cadena complementaria desplazada. La extensión del cebador inverso en una reacción de secuenciación por síntesis genera una lectura de secuenciación de cadena inversa.

En una variación, se genera una lectura de secuenciación directa iniciada a partir de un cebador convencional que se une a un sitio de unión del cebador directo en un adaptador de una molécula diana monocatenaria flanqueada por un adaptador, tal como se muestra en la figura 1 o en la figura 10. La cadena complementaria generada por la lectura de secuenciación directa se desplaza entonces restaurando la molécula diana monocatenaria original, como se muestra en la figura 1 o en la figura 10. A continuación, el método procede como se ha descrito anteriormente con el cebador

de tallo-bucle o el cebador funcional doble, excepto que la extensión para la región de unión a la diana que forma dúplex con el sitio de unión del cebador directo es una extensión que no genera una lectura de secuenciación (a veces denominada extensión oscura).

REIVINDICACIONES

1. Un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana, que comprende:

- 5 (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprenda una región de unión a la diana 3' y un tallo-bucle 5' para formar una mezcla de reacción, en donde la región de unión a la diana se une al adaptador 3';
- 10 (b) ligar un extremo 5' libre del tallo-bucle al adaptador 3';
- (c) realizar una reacción de extensión, en donde la región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y, opcionalmente, proporciona una lectura de secuenciación del ácido nucleico diana;
- (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle;
- 15 (e) realizar una reacción de extensión, en donde el cebador de bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, y la primera cadena complementaria se desplaza del dúplex en formación con el ácido nucleico diana;
- (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la primera cadena complementaria desplazada; y
- 20 (g) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la primera cadena complementaria desplazada y proporciona una lectura de secuenciación de la primera cadena complementaria desplazada.

2. El método de la reivindicación 1:

- 25 (i) que comprende además, antes de la etapa (a), poner en contacto el ácido nucleico diana monocatenario con un segundo cebador directo que comprende una región de unión a la diana, que se une a un sitio complementario en el adaptador 3';
- realizar una reacción de secuenciación por extensión en donde la región de unión a la diana del segundo cebador
- 30 directo inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y
- desplazar la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana;
- (ii) en donde la etapa (c) es una reacción de secuenciación por extensión que proporciona una lectura de secuencia del ácido nucleico diana; o
- 35 (iii) en donde la etapa (c) se realiza antes de la etapa (b).

3. Un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana, que comprende

- 40 (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprende una región de unión a la diana 3' y un tallo-bucle 5' para formar una mezcla de reacción, en donde la región de unión a la diana se une al adaptador 3' y en donde la región de unión a la diana comprende un sitio escindible;
- (b) ligar un extremo 5' libre del tallo-bucle al adaptador 3';
- 45 (c) realizar una reacción de secuenciación por extensión, en donde la región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; escindir el sitio escindible, y realizar otra reacción de extensión en donde el resto de la región de unión a la diana después de la escisión inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana;
- (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle;
- 50 (e) realizar una reacción de extensión en donde el cebador de bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana;
- (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la cadena complementaria adicional desplazada;
- 55 (g) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la cadena complementaria adicional desplazada y proporciona una lectura de secuenciación de la cadena complementaria adicional desplazada.

4. Un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana, que comprende

- 60 (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo en horquilla que comprende una región de unión a la diana 3' y un tallo-bucle 5' y un cebador directo lineal para formar una mezcla
- 65 de reacción, en donde la región de unión a la diana del cebador directo de horquilla-tallo y el cebador directo lineal unen el adaptador 3' con el cebador de tallo-bucle entre el cebador lineal y el extremo 3';

- (b) ligar un extremo 5' libre del tallo-bucle con el extremo 3' de la molécula de adaptador en el extremo 3' del ácido nucleico diana;
- (c) realizar una reacción de secuenciación por extensión, en donde el cebador lineal inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y realizar otra reacción de extensión en donde la región de unión a la diana del cebador directo de tallo-bucle inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana;
- (d) poner en contacto la mezcla de reacción con un cebador de bucle que hibrida con la región de bucle del tallo-bucle;
- (e) realizar una reacción de extensión en donde el cebador de bucle inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana;
- (f) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la cadena complementaria adicional desplazada;
- (g) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la cadena complementaria adicional desplazada y proporciona una lectura de secuenciación de la cadena complementaria adicional desplazada, opcionalmente, en donde el cebador directo de tallo-bucle contiene un nucleótido bloqueado para la extensión y el método comprende además eliminar el bloqueo antes de extender el cebador directo de tallo-bucle.

5. Un método para obtener lecturas de secuenciación por pares de un ácido nucleico diana, que comprende

- (a) poner en contacto un ácido nucleico diana monocatenario flanqueado en su extremo 5' por un adaptador 5' inmovilizado en un soporte y en su extremo 3' por un adaptador 3' con un cebador directo que comprende una primera y segunda regiones de unión a la diana 3' que flanquean una región intermedia para formar una mezcla de reacción, en donde la primera y la segunda regiones de unión a la diana se unen a sitios complementarios en la molécula de adaptador 3', estando los respectivos sitios en posición distal y proximal al extremo 3' del adaptador 3';
- (b) realizar una reacción de extensión, en donde la primera región de unión a la diana inicia la síntesis de una primera cadena complementaria del ácido nucleico diana y, opcionalmente, proporciona una lectura de secuenciación del ácido nucleico diana;
- (c) realizar una segunda reacción de extensión en donde la segunda región de unión a la diana inicia la síntesis de una segunda cadena complementaria del ácido nucleico diana, que desplaza la cadena complementaria del dúplex en formación con el ácido nucleico diana;
- (d) poner en contacto la mezcla de reacción con un cebador inverso, que hibrida con un complemento del adaptador 5' en el extremo 3' de la primera cadena complementaria desplazada; y
- (e) realizar una reacción de secuenciación por extensión, en donde el cebador inverso inicia la síntesis de una cadena complementaria de la primera cadena complementaria desplazada y proporciona una lectura de secuenciación de la primera cadena complementaria desplazada.

6. El método de la reivindicación 5:

- (i) que comprende además, antes de la etapa (a), poner en contacto el ácido nucleico diana monocatenario con un segundo cebador directo que comprende una región de unión a la diana, que se une a un sitio complementario de la molécula de adaptador en el extremo 3' del ácido nucleico diana;
- realizar una reacción de secuenciación por extensión en donde la región de unión a la diana del segundo cebador directo inicia la síntesis de una cadena complementaria adicional del ácido nucleico diana y proporciona una lectura de secuenciación del ácido nucleico diana; y
- desplazar la cadena complementaria adicional del dúplex en formación con el ácido nucleico diana;
- (ii) en donde la región intermedia comprende un nucleótido invertido;
- (iii) en donde la región intermedia comprende un enlace no nucleotídico;
- (iv) en donde la segunda región de unión a la diana está bloqueada para la extensión y el método comprende además eliminar el bloqueo para la extensión antes de realizar la extensión iniciada por la segunda región de unión a la diana, opcionalmente, en donde la segunda región de unión a la diana está bloqueada para la extensión por un grupo fosfato terminal 3' y eliminar el bloqueo que comprende el tratamiento con una fosfatasa para eliminar el fosfato;
- (v) en donde la segunda extensión se realiza con una polimerasa con actividad de desplazamiento de cadena; o
- (vi) en donde el adaptador 3' comprende además un marcador de muestra entre los sitios complementarios para la primera y segunda regiones de unión a la diana.

7. El método de cualquier reivindicación anterior, en donde el adaptador 5' o 3', o ambos, comprende además un marcador molecular y/o un marcador de muestra.

8. El método de cualquier reivindicación anterior, en donde el ácido nucleico monocatenario flanqueado por los adaptadores 5' y 3' se produce por amplificación en emulsión, opcionalmente:

- (i) que comprende además romper la emulsión antes de realizar la etapa (a);

- (ii) en donde el soporte en el que se inmoviliza el ácido nucleico diana monocatenario se utilizó en PCR en emulsión para generar el ácido nucleico diana monocatenario flanqueado por adaptadores 5' y 3'; o
- (iii) en donde el soporte en el que se inmoviliza el ácido nucleico diana monocatenario se fijó después de la PCR en emulsión.

- 5 9. El método de cualquier reivindicación anterior, en donde el ácido nucleico diana monocatenario inmovilizado en un soporte forma parte de una población clonal de dichos ácidos nucleicos diana inmovilizados en el mismo soporte, opcionalmente, en donde el soporte es una perla o una región direccionable dentro de una matriz.
- 10 10. El método de cualquier reivindicación anterior, en donde la secuenciación es mediante: (i) secuenciación 454, de Ion Torrent o de Ultima; y/o (ii) secuenciación de molécula única o secuenciación clonal.
- 15 11. El método de cualquier reivindicación anterior, en donde la longitud del ácido nucleico diana supera la longitud máxima de lectura del método de secuenciación, opcionalmente, en donde el ácido nucleico diana tiene aproximadamente 170 nucleótidos y la longitud máxima de lectura del método de secuenciación es de aproximadamente 150 nucleótidos del ácido nucleico diana.
- 20 12. El método de cualquier reivindicación anterior:
 - (i) en donde las lecturas de secuenciación tienen una región no solapada; y/o
 - (ii) que comprende además determinar una lectura de secuenciación compuesta a partir de las lecturas de secuenciación directa e inversa en donde cualquier posición de discordancia entre las lecturas de secuenciación directa e inversa se deja abierta o como nucleótidos alternativos que ocupan la posición en las lecturas de secuenciación directa e inversa.
- 25 13. Un cebador que comprende una primera y una segunda regiones de unión a la diana, cada una con un extremo 3' libre separadas por un nucleótido invertido.
- 30 14. Un kit que comprende un cebador con una primera y una segunda regiones de unión a la diana, cada una con un extremo 3' libre separadas por un nucleótido invertido, primer y segundo adaptadores, o un adaptador único con regiones monocatenarias asimétricas dispuestas para flanquear cadenas individuales de ácido nucleico diana bicatenario con el primer y segundo adaptadores y un segundo cebador, en donde la primera y la segunda regiones de unión a la diana se unen al primer adaptador y el segundo cebador se une al segundo adaptador, opcionalmente, en donde uno de los extremos libres 3' está bloqueado con un grupo fosfato y el kit comprende además una fosfatasa para eliminar el fosfato.
- 35 15. El kit de la reivindicación 14 que comprende además una polimerasa con actividad de desplazamiento de cadena.

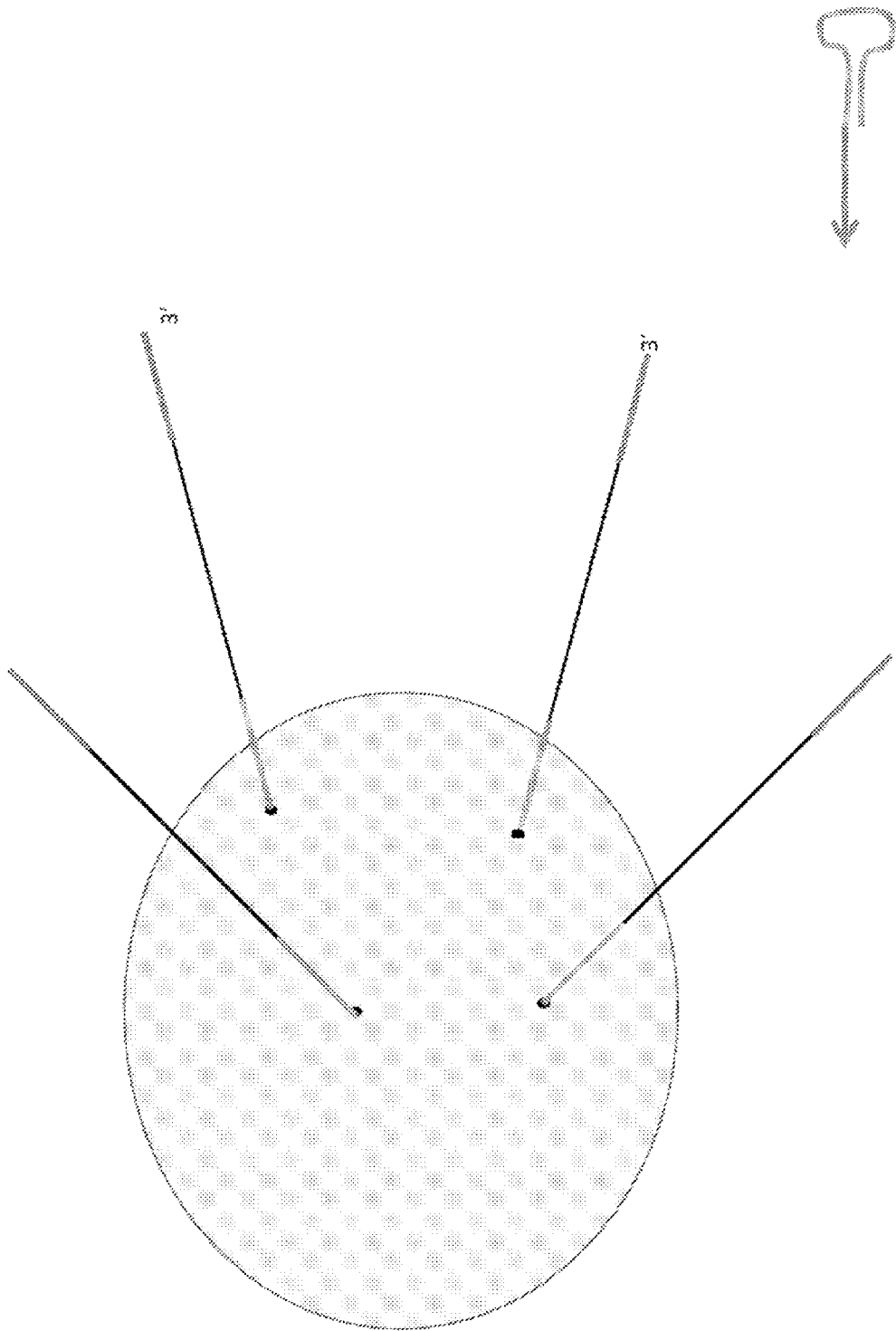


Fig. 1

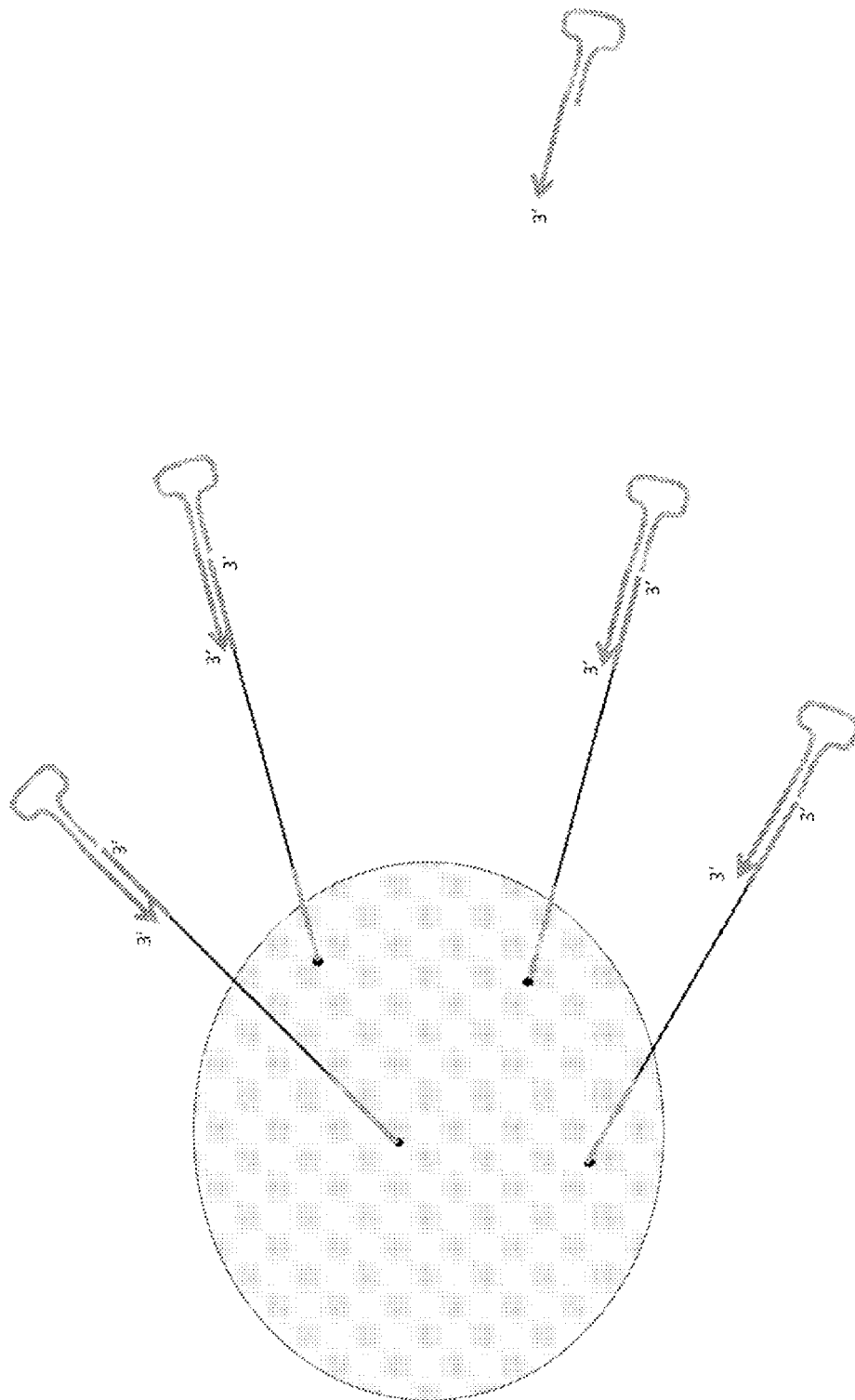
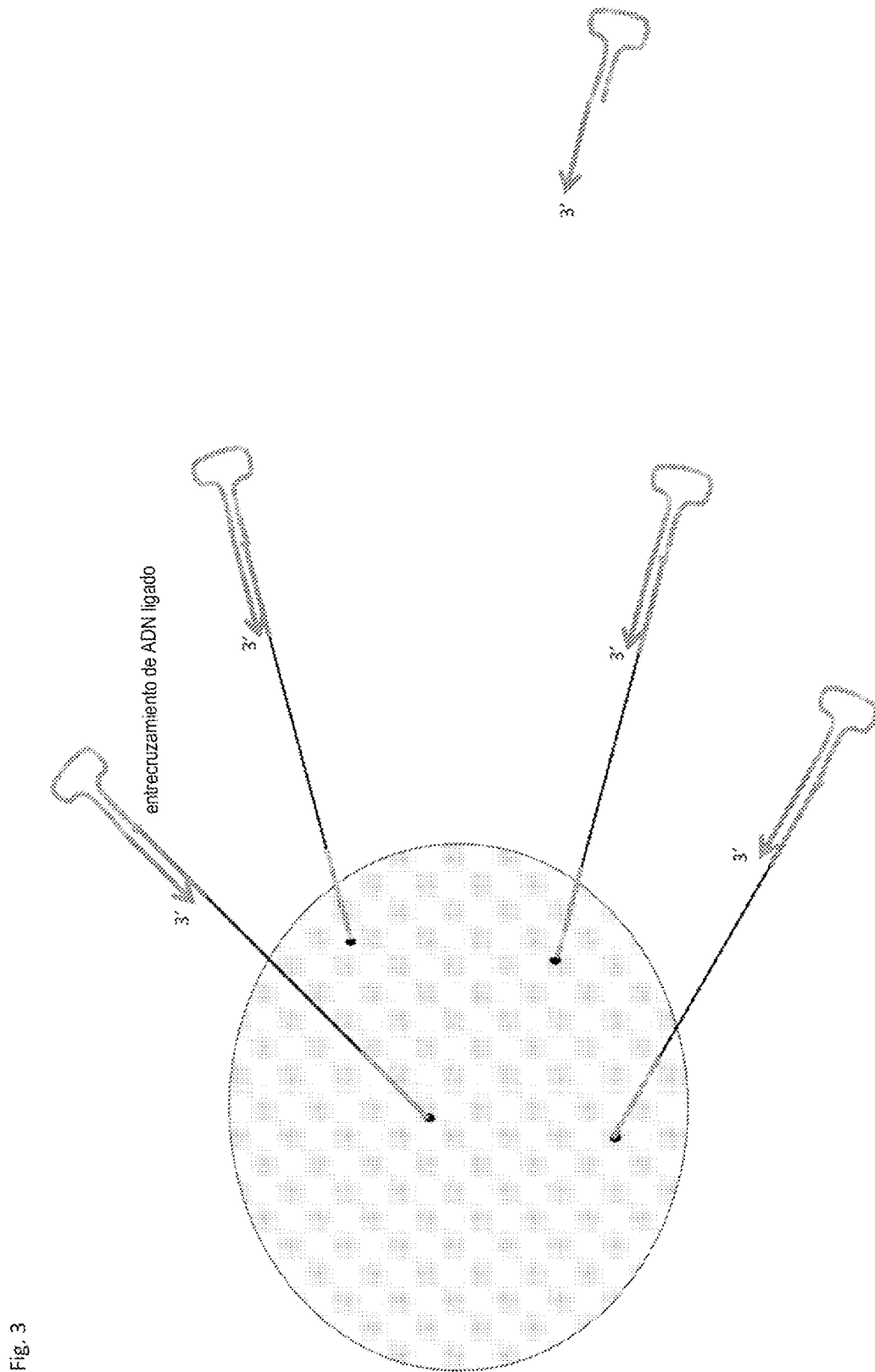


Fig. 2



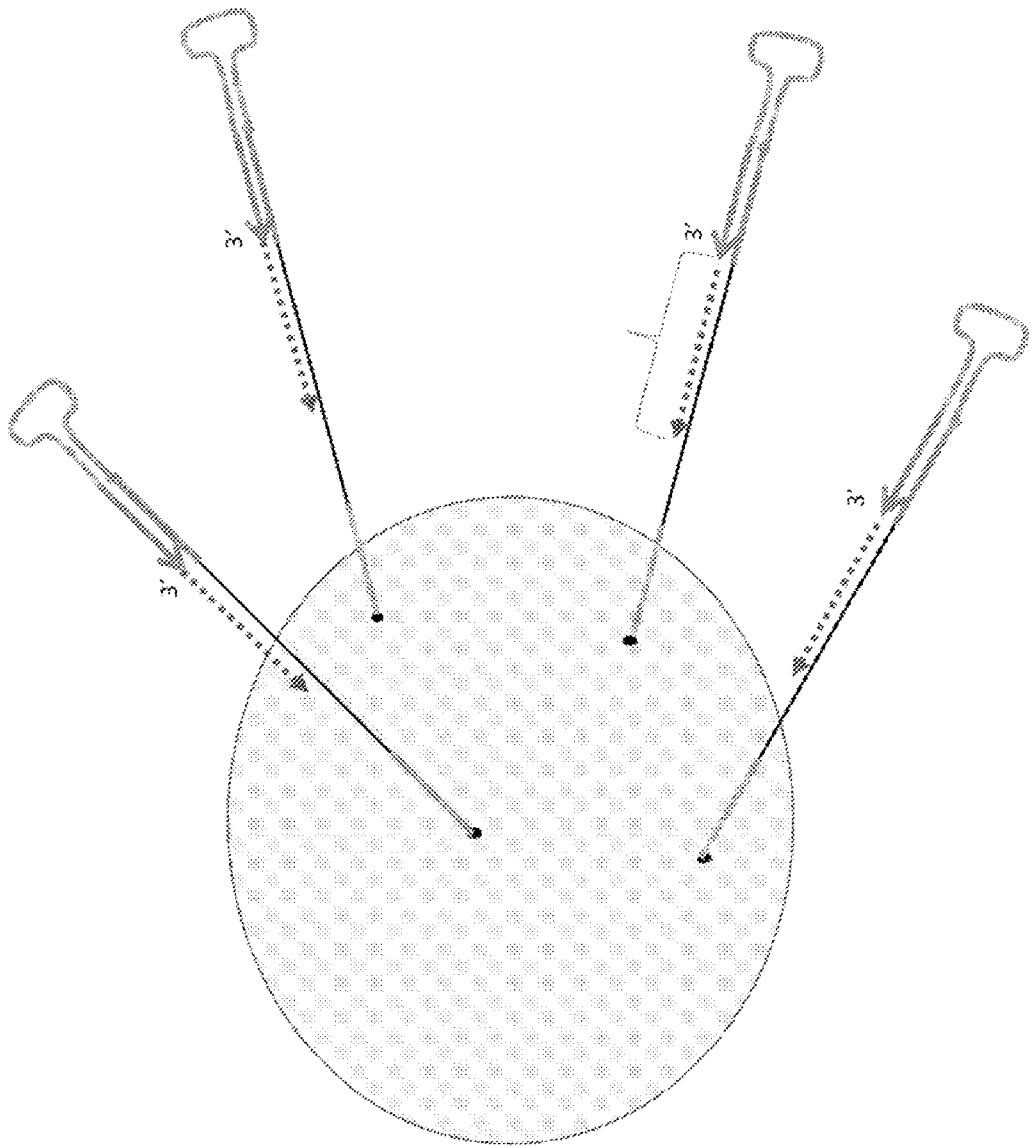


Fig. 4

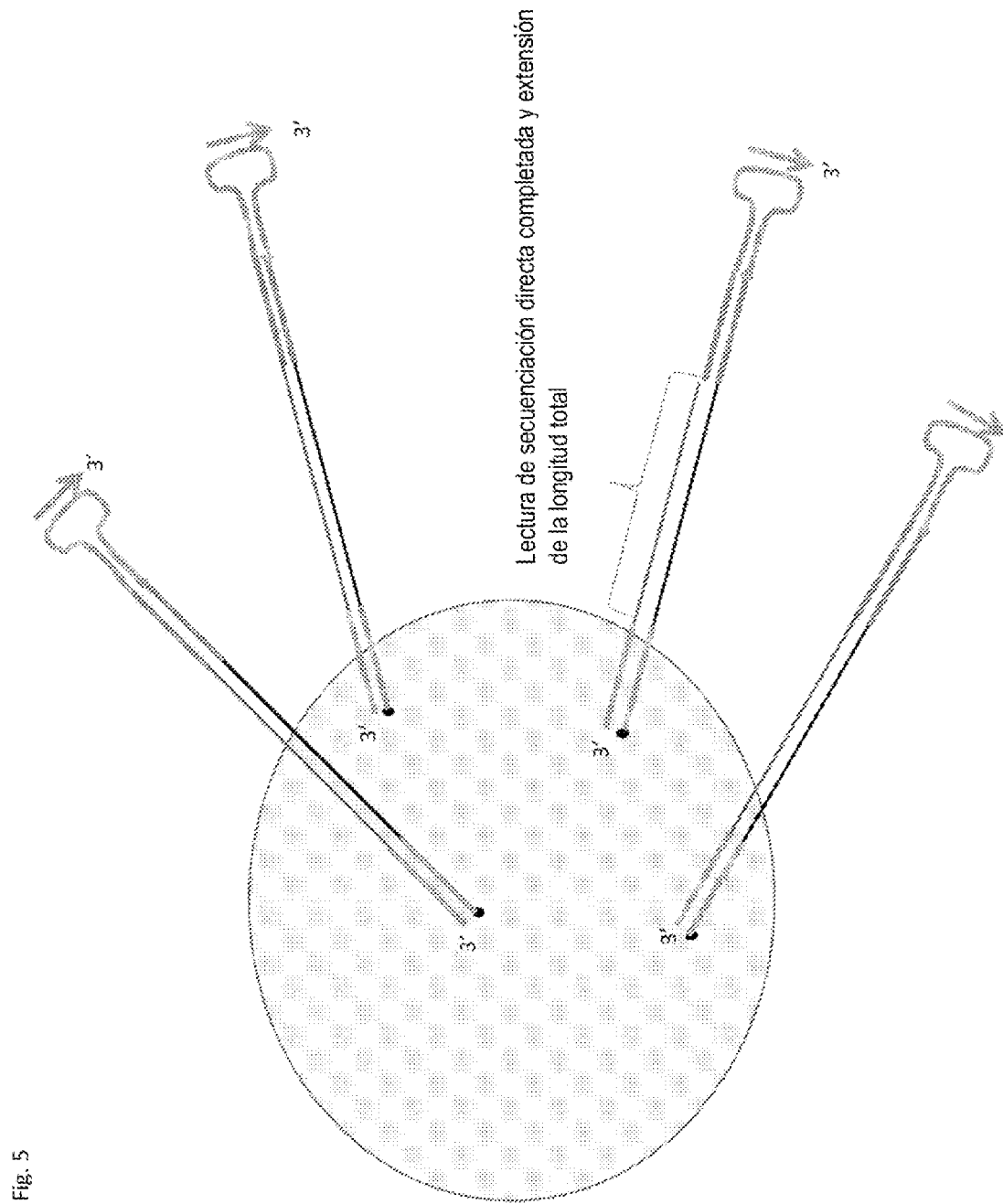
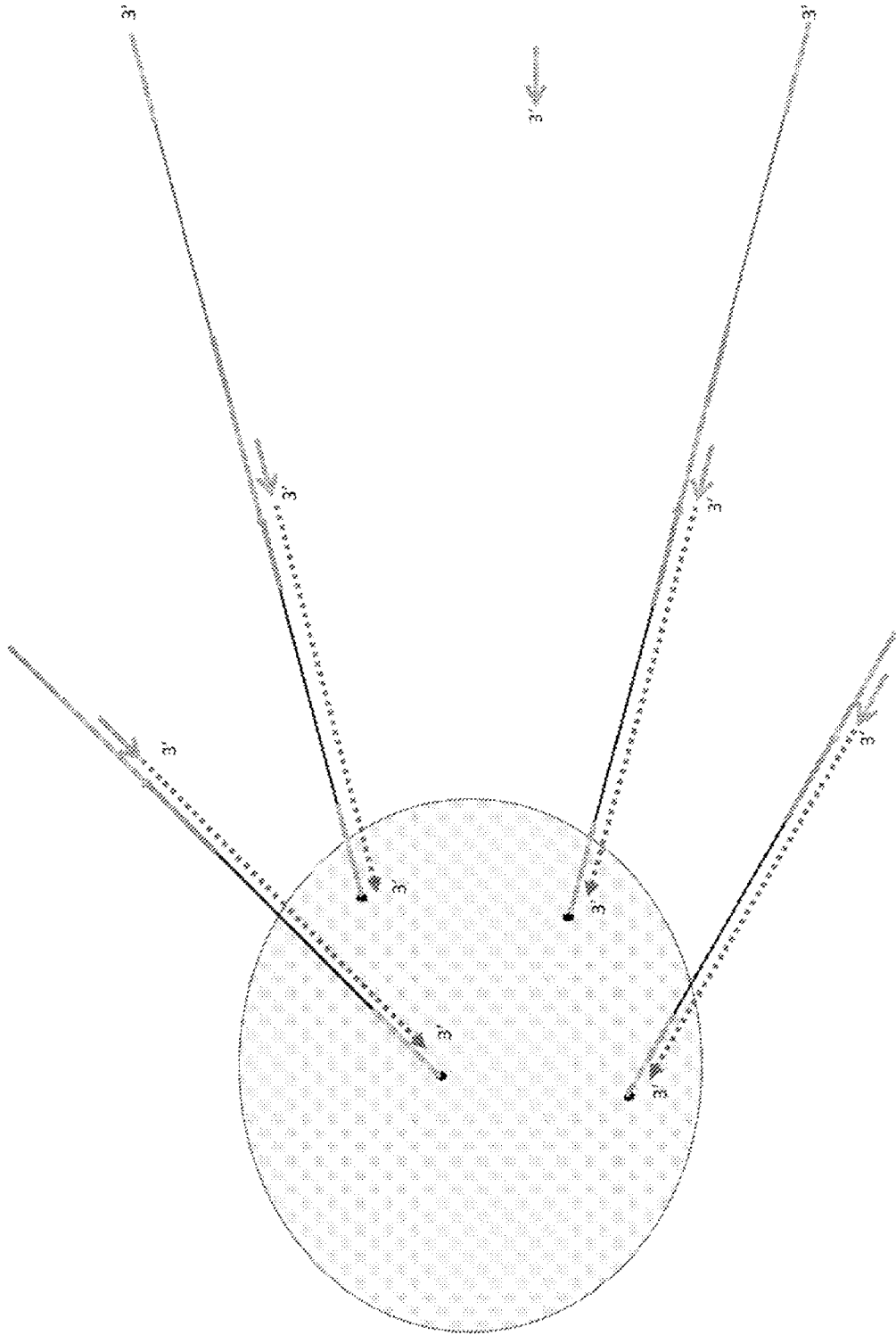


Fig. 6



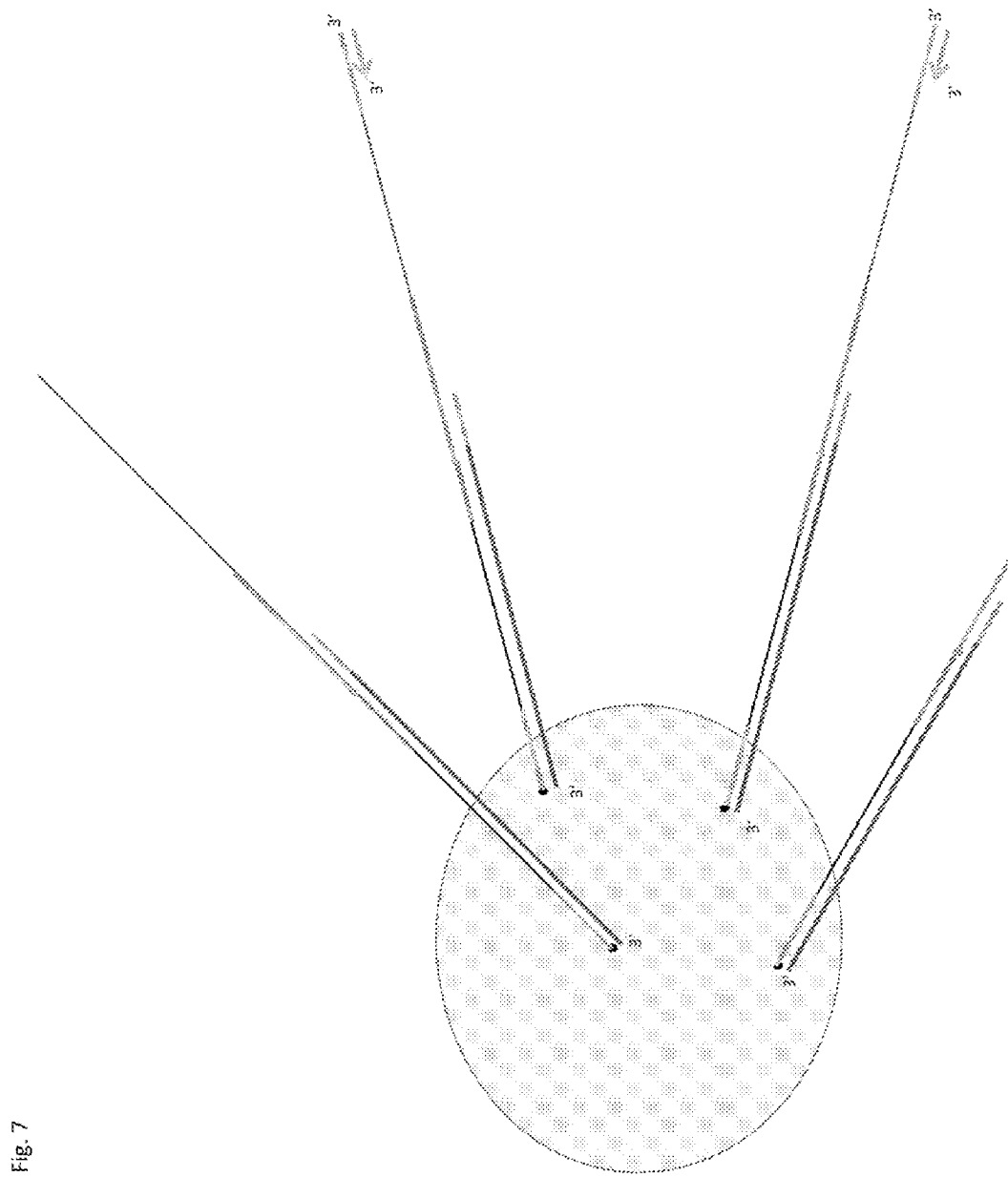


Fig. 7

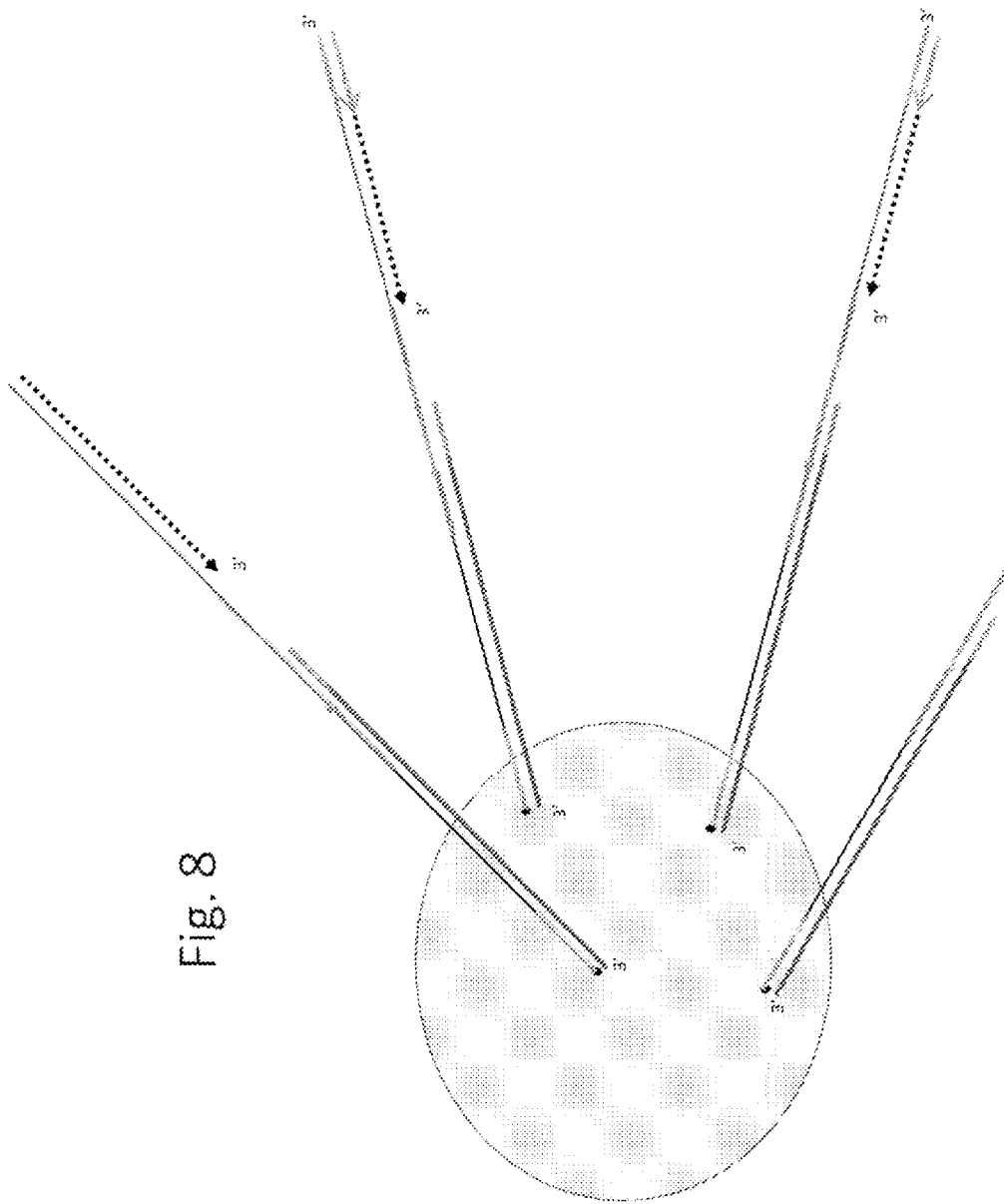
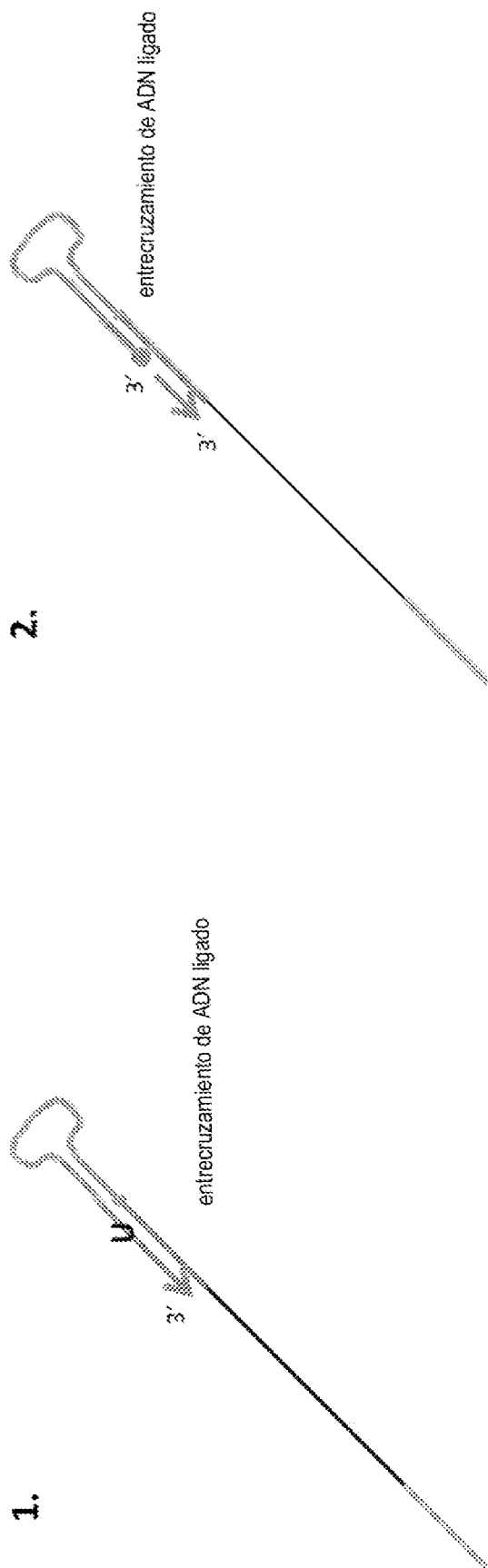


Fig. 9



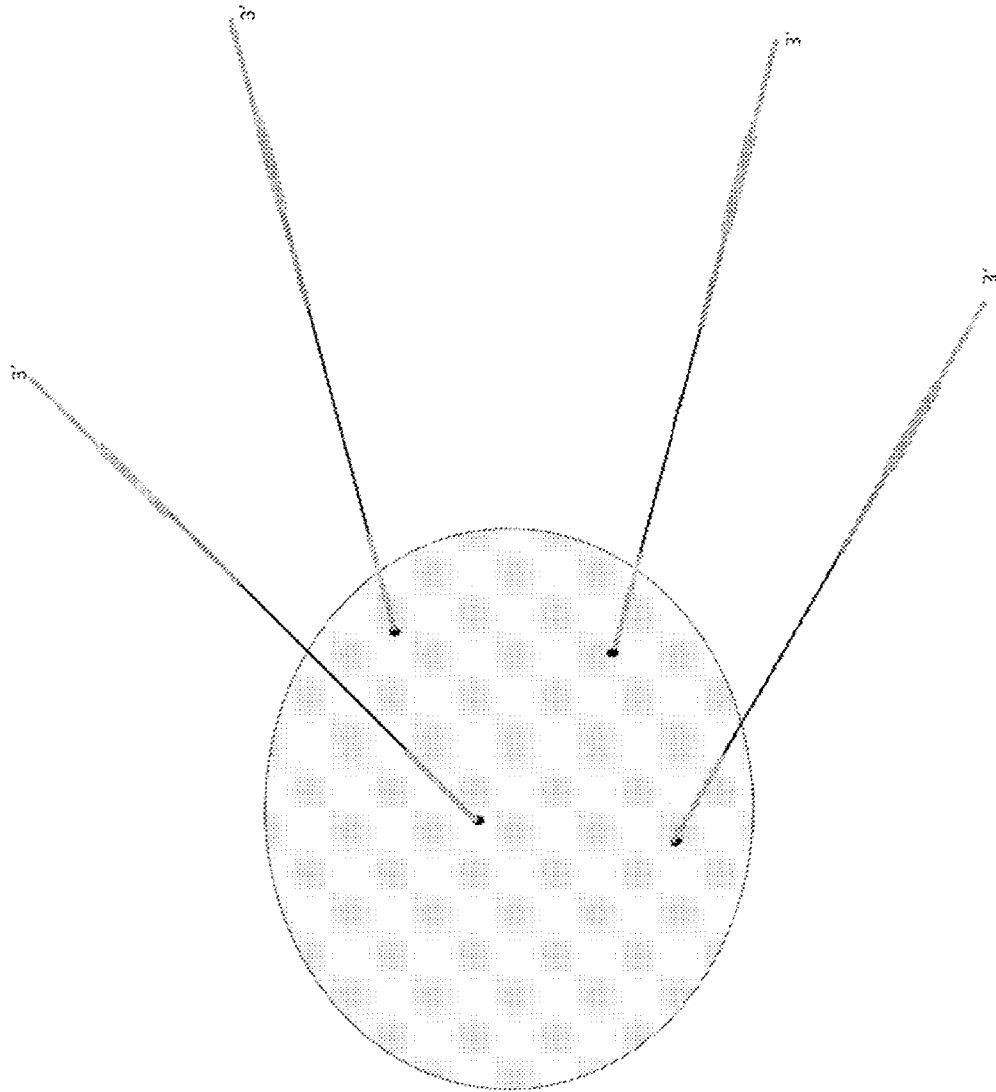
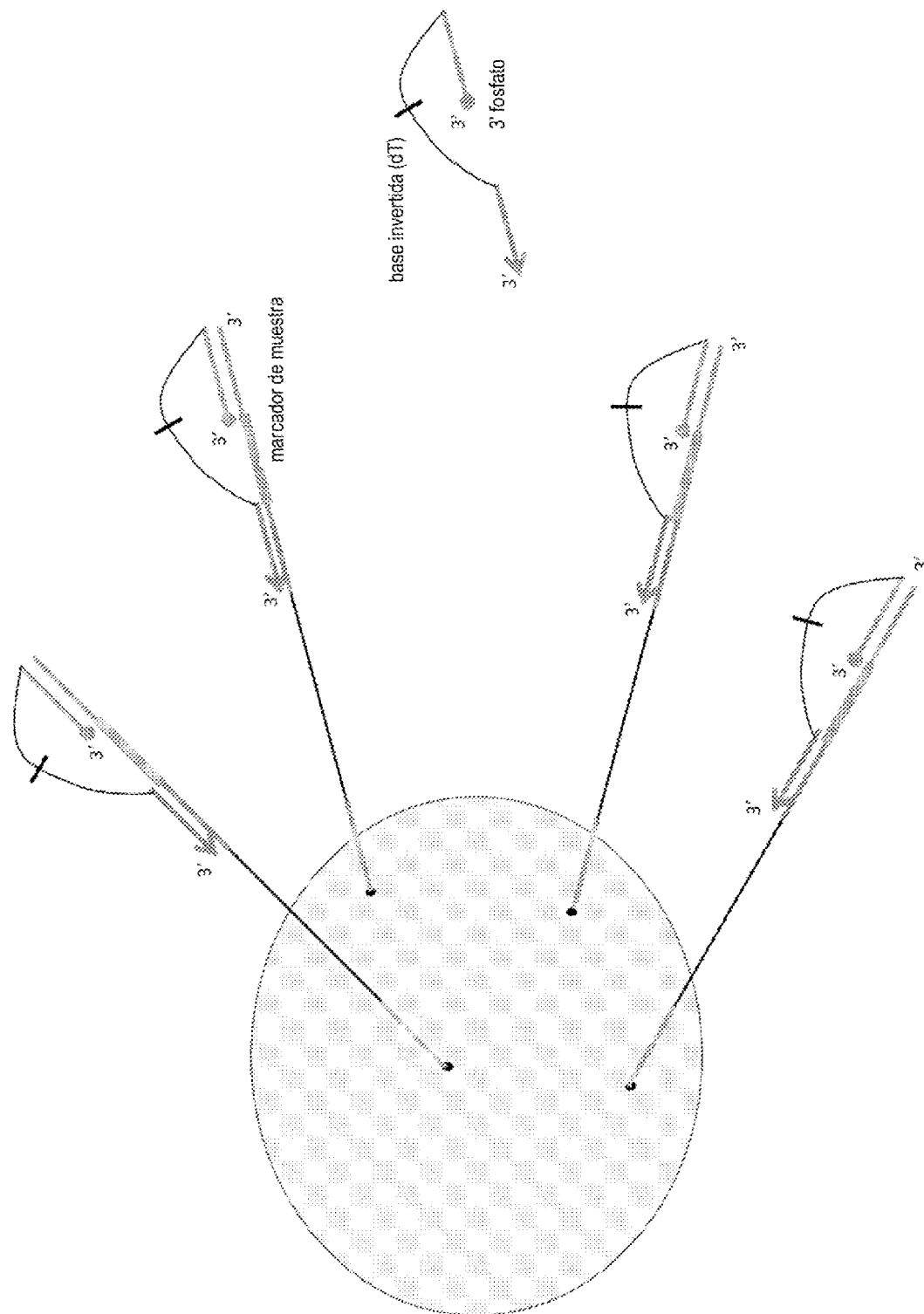


Fig. 10

Fig. 11



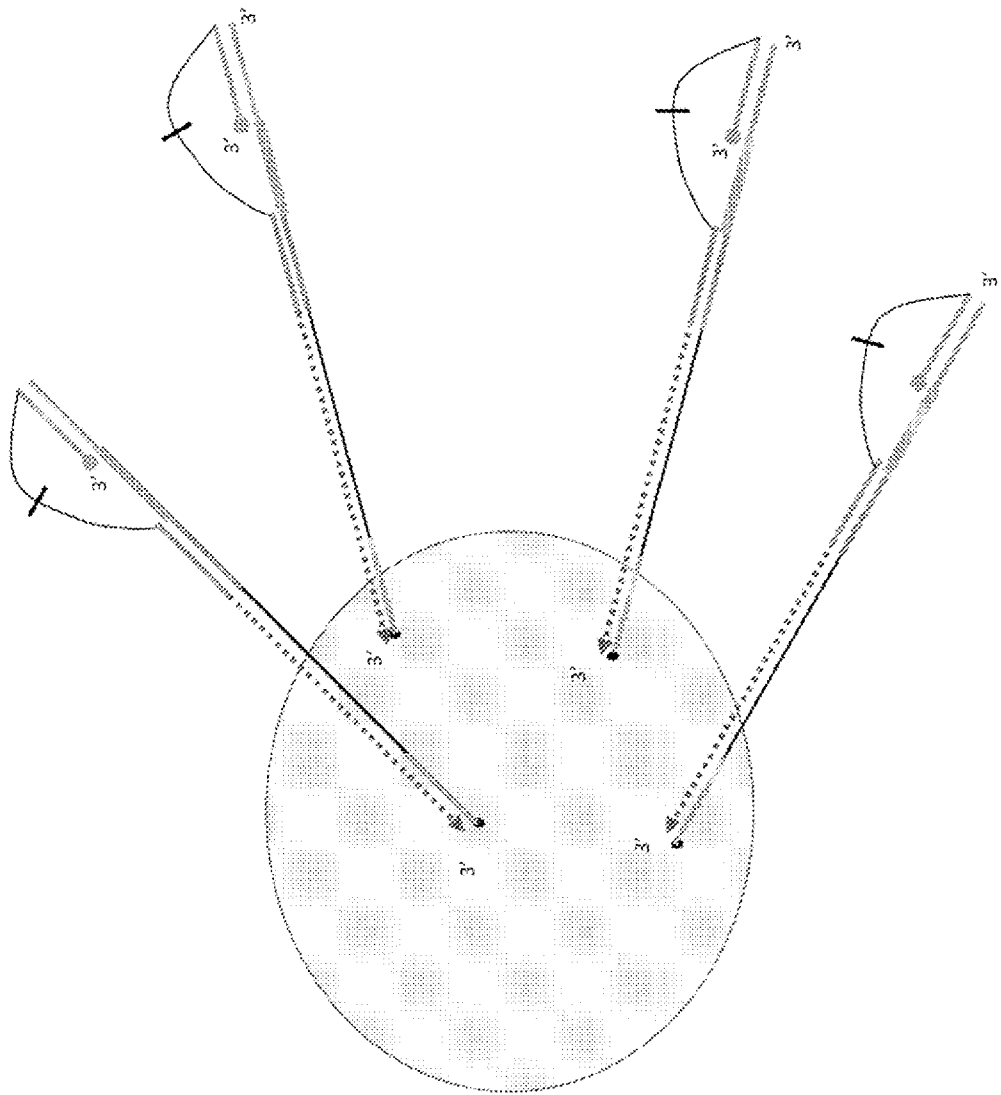


Fig. 12

Fig. 13

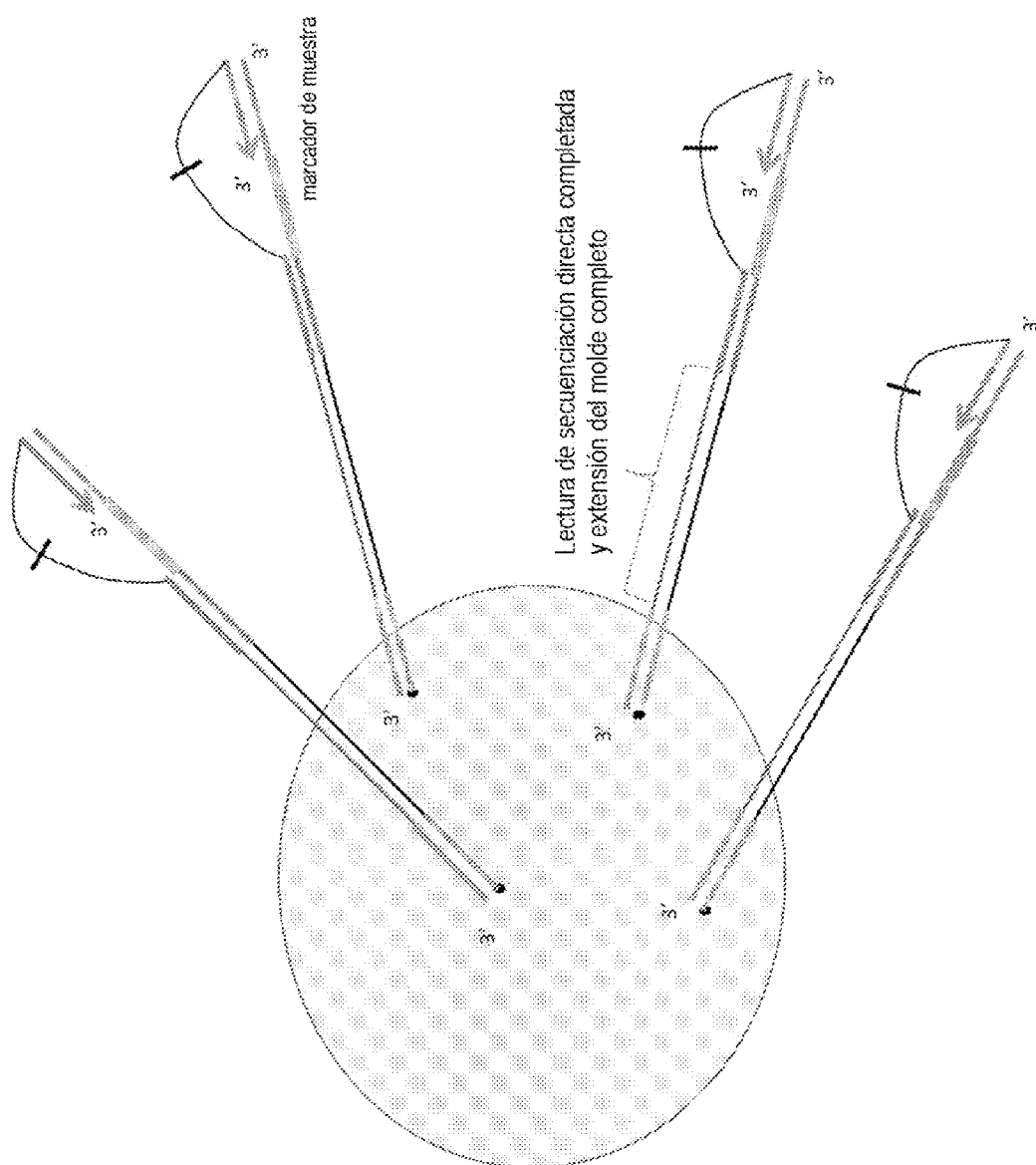


Fig. 14

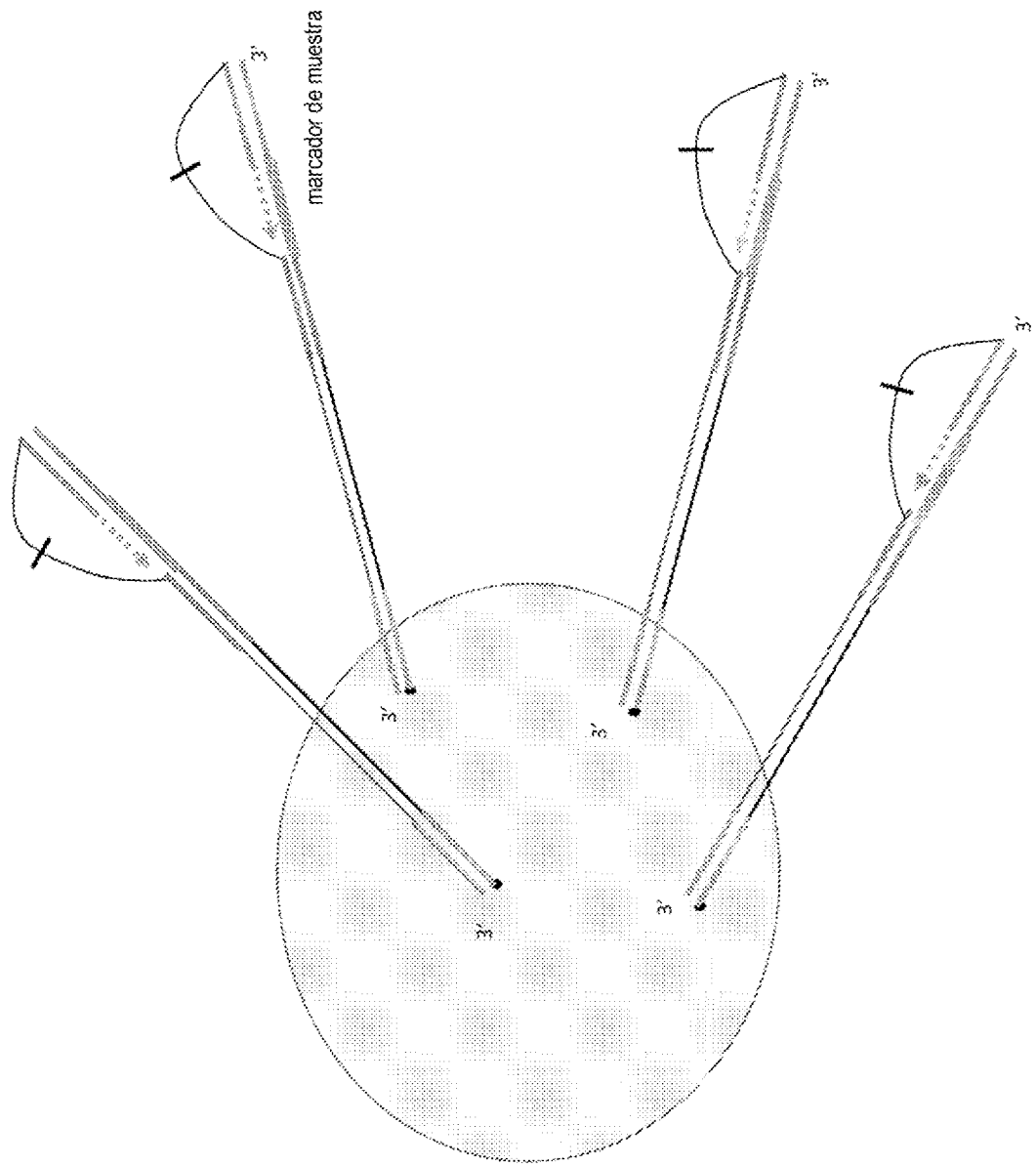
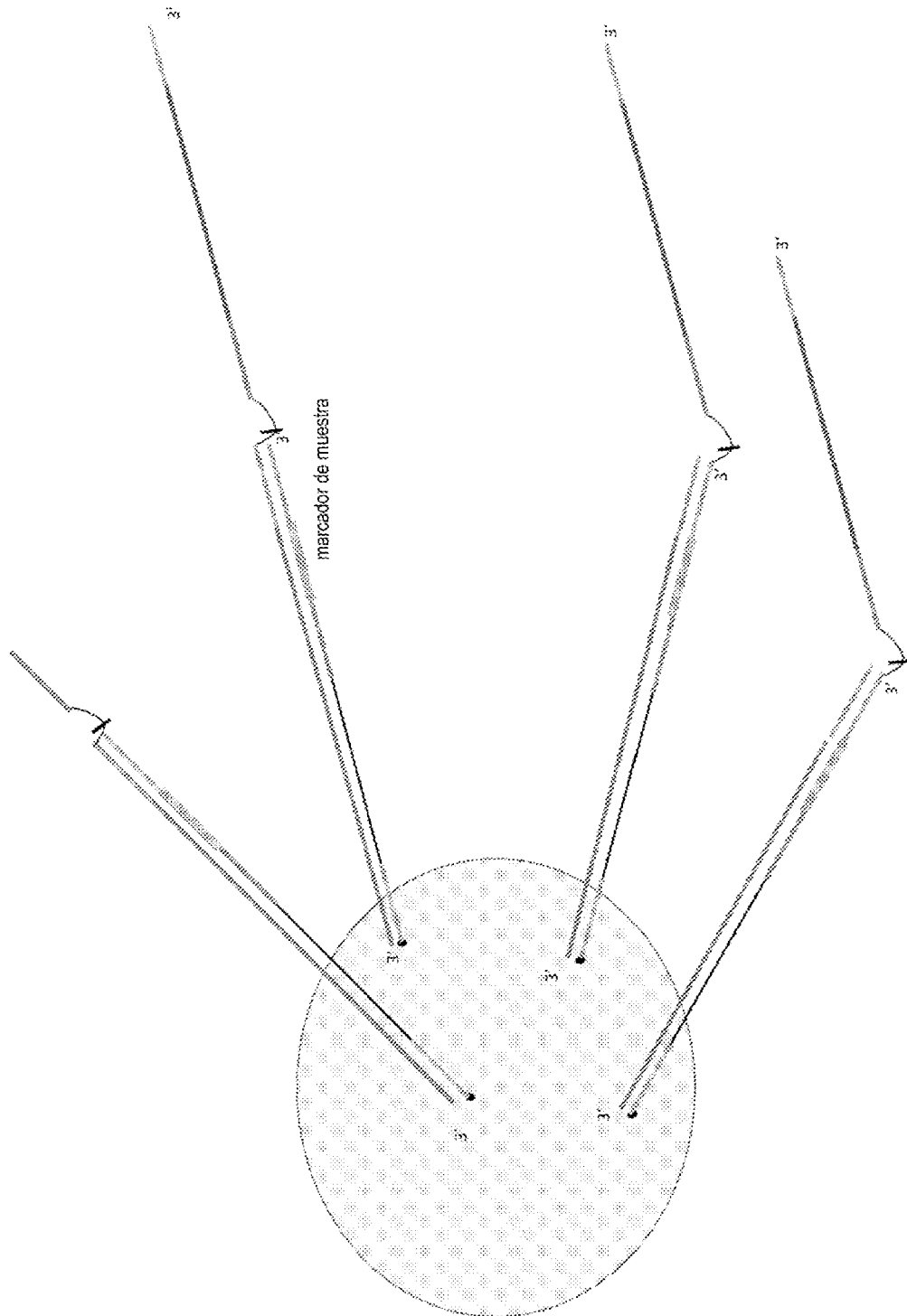


Fig. 15



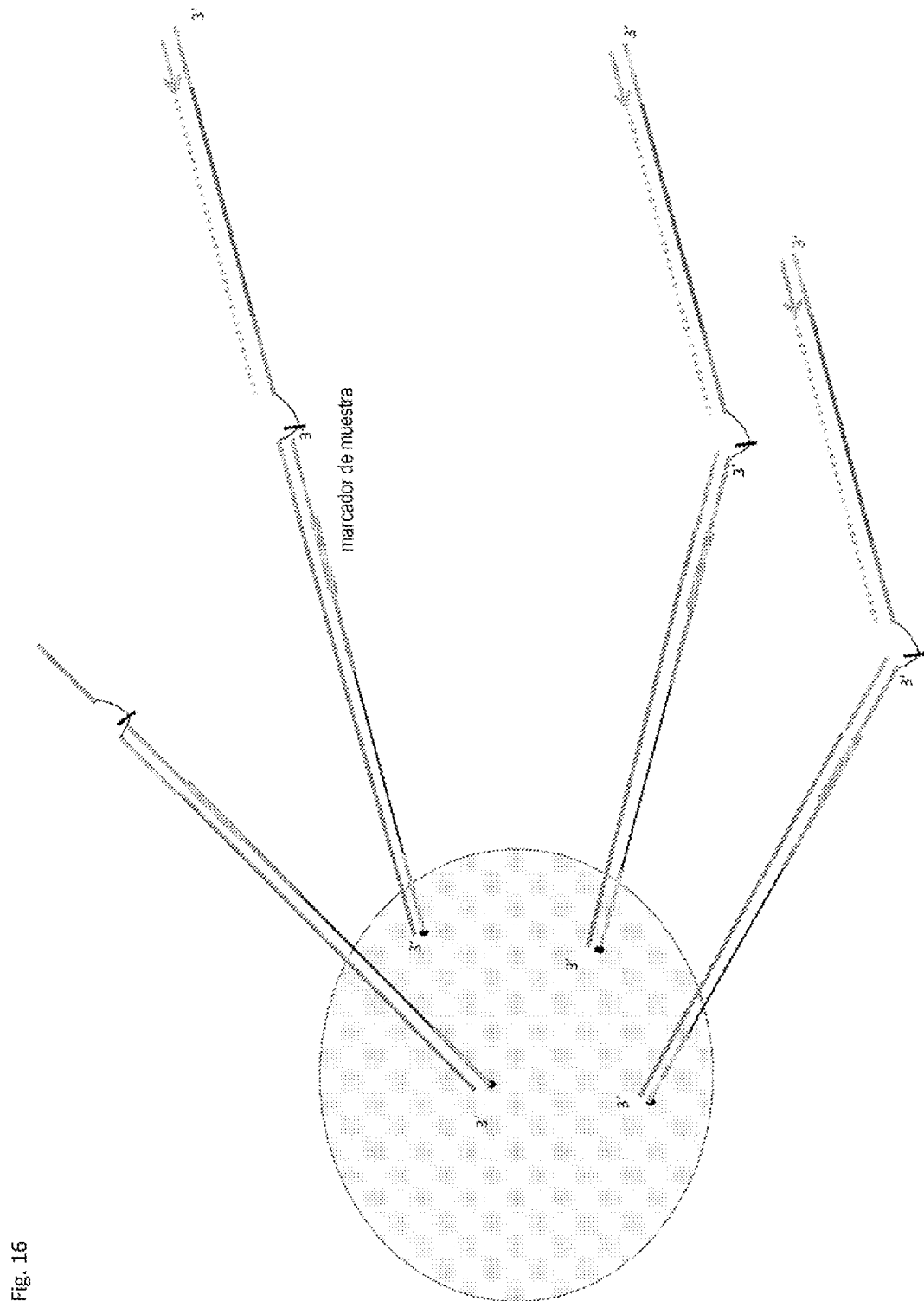


Fig. 16

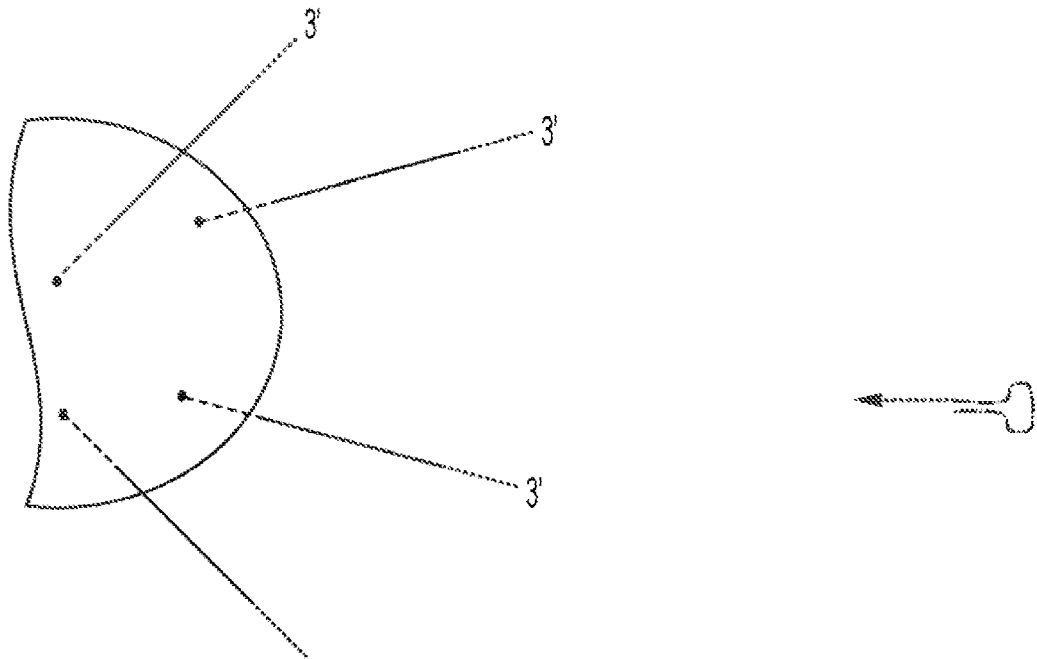


FIG. 1

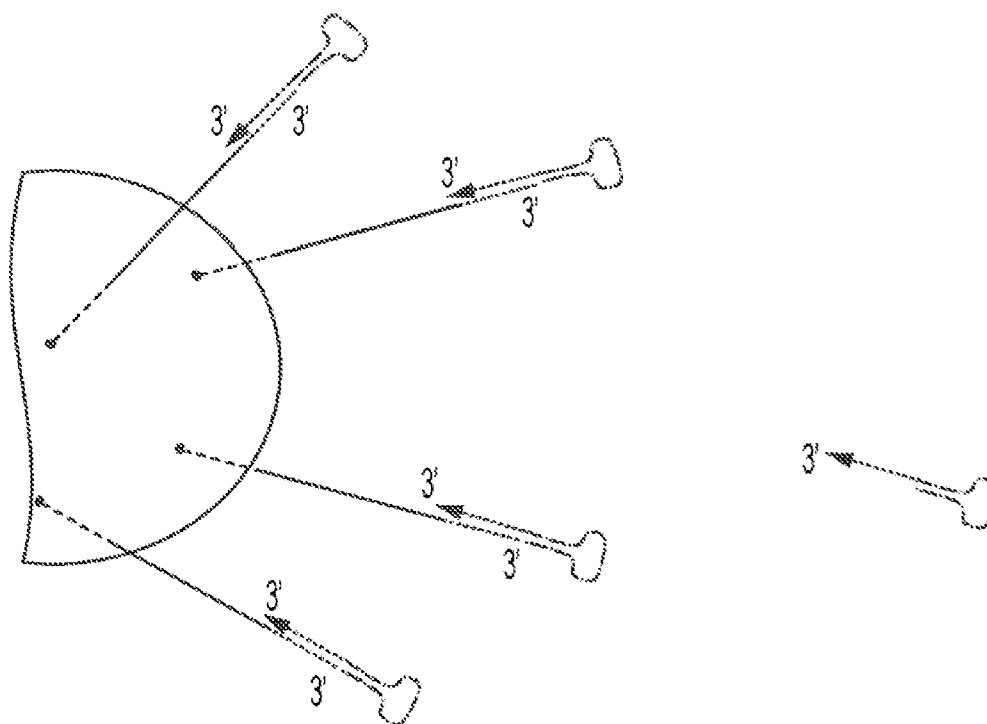


FIG. 2

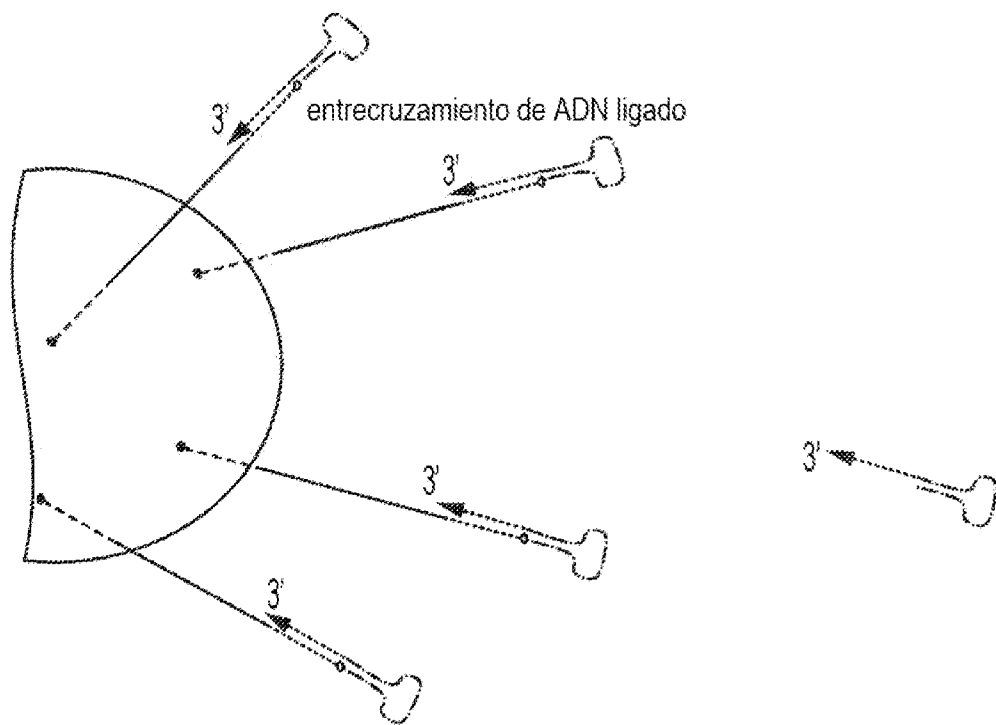


FIG. 3

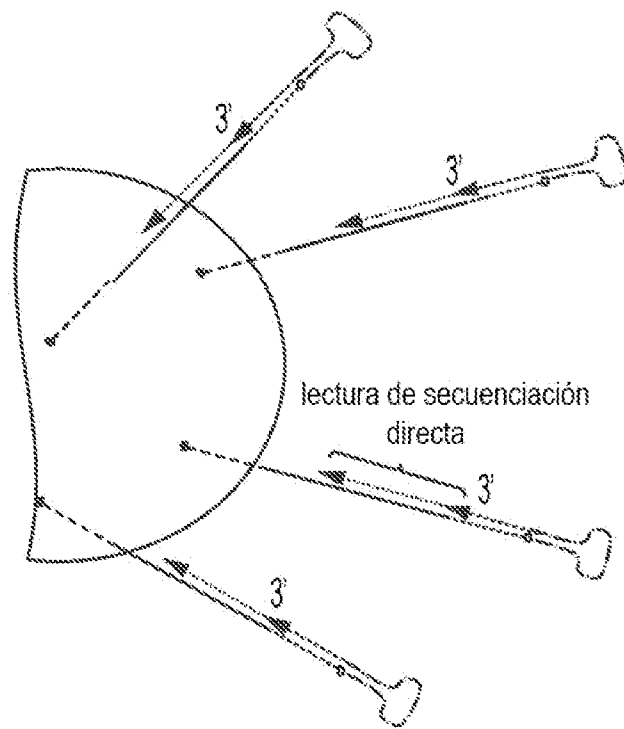


FIG. 4

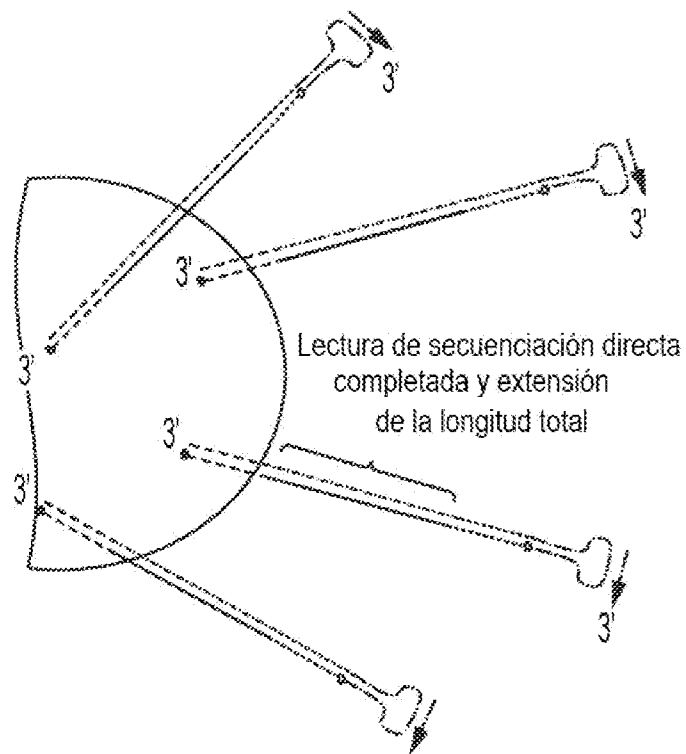


FIG. 5

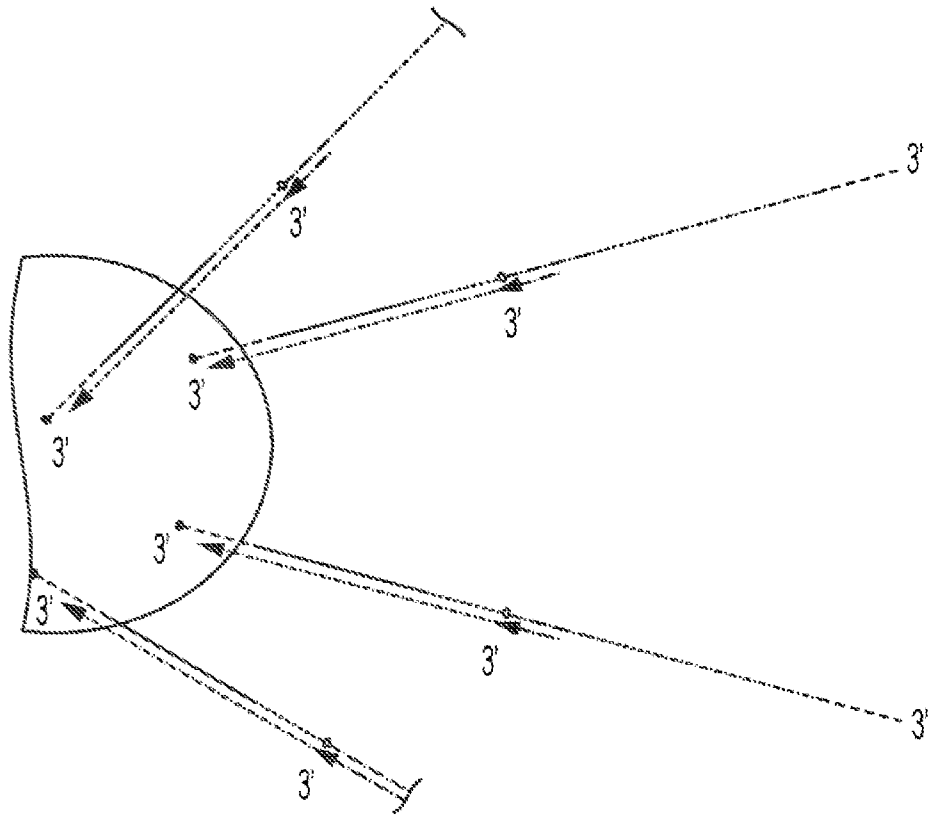


FIG. 6

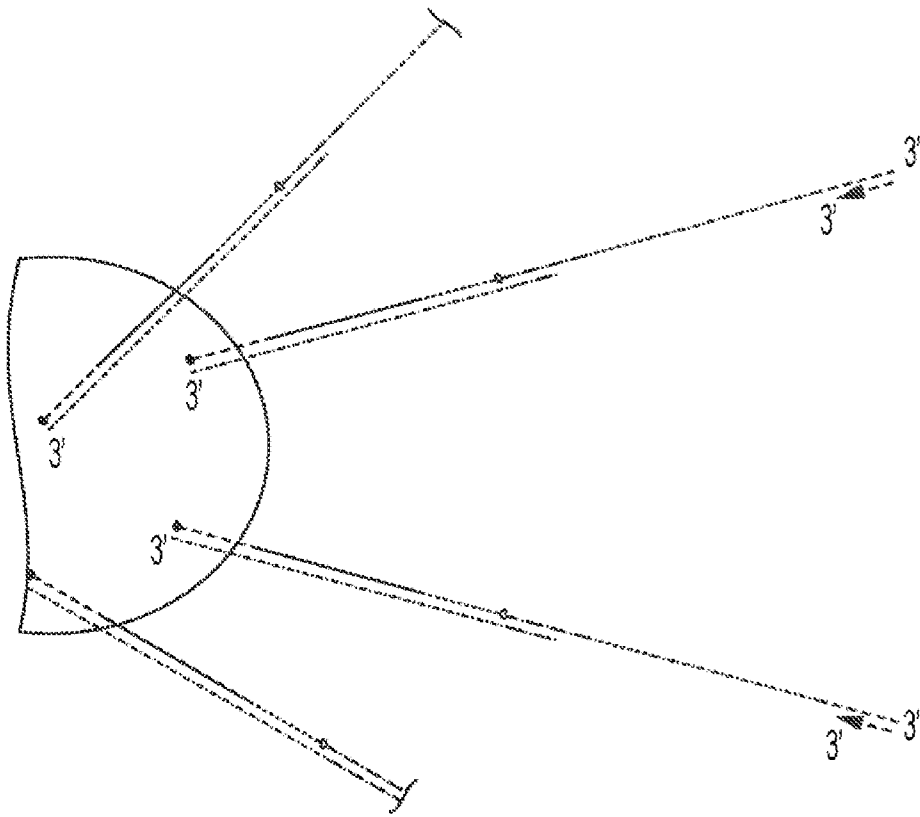


FIG. 7

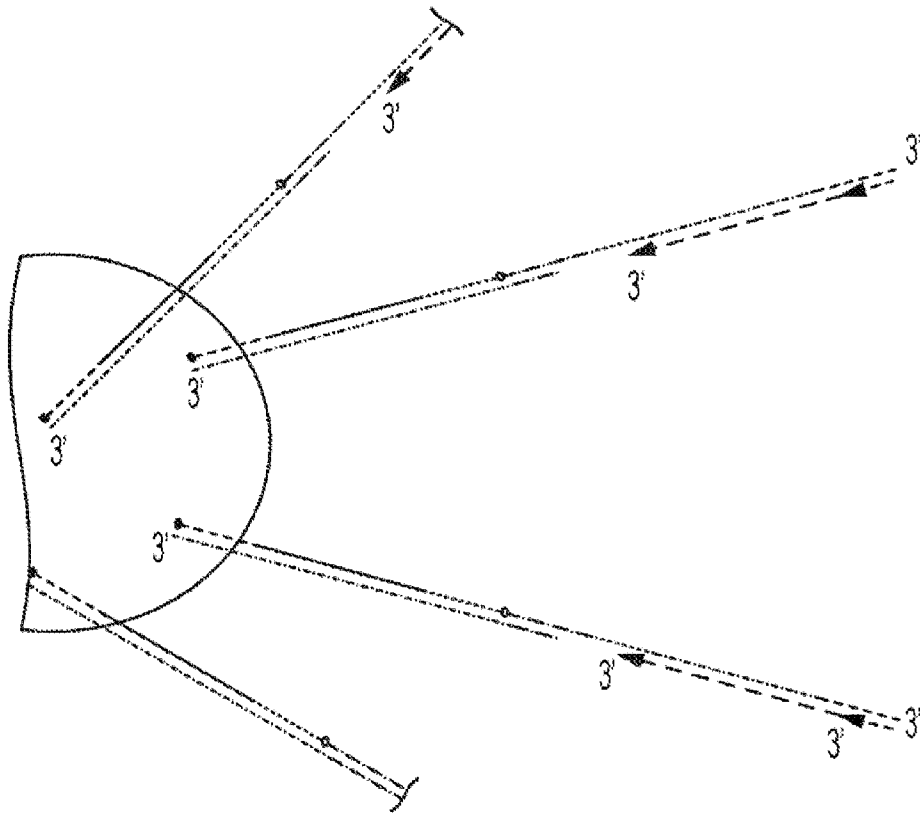


FIG. 8

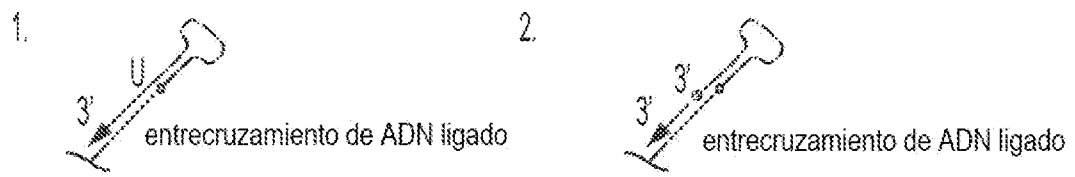


FIG. 9

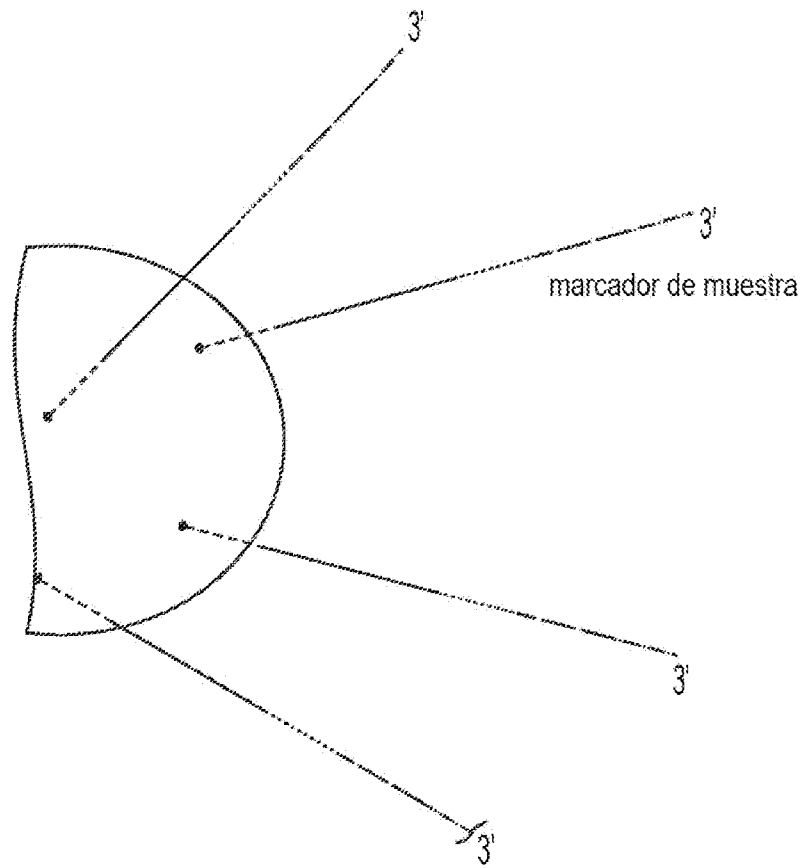


FIG. 10

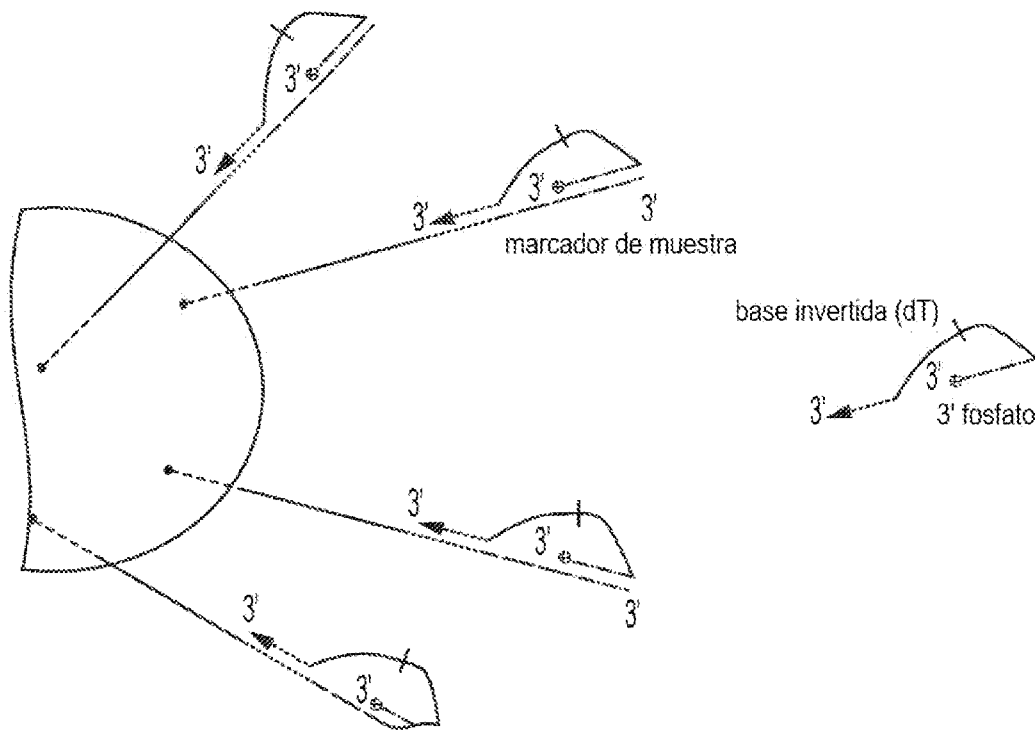


FIG. 11

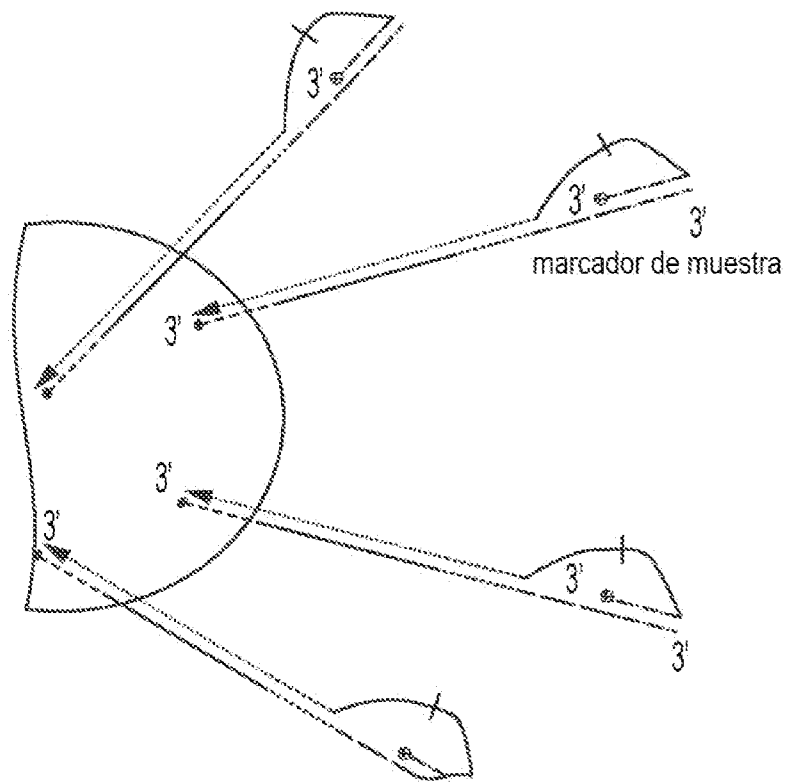


FIG. 12

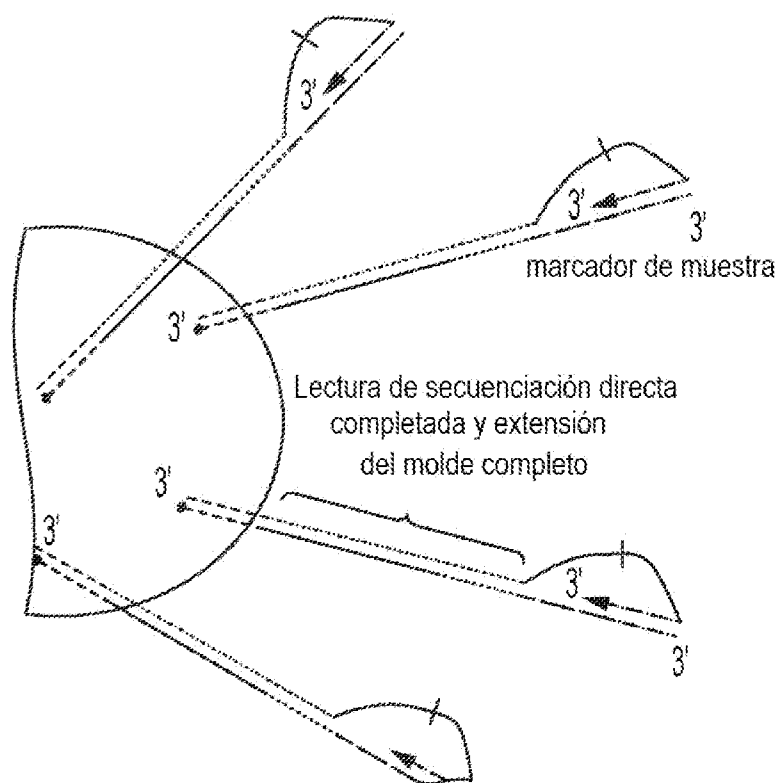


FIG. 13

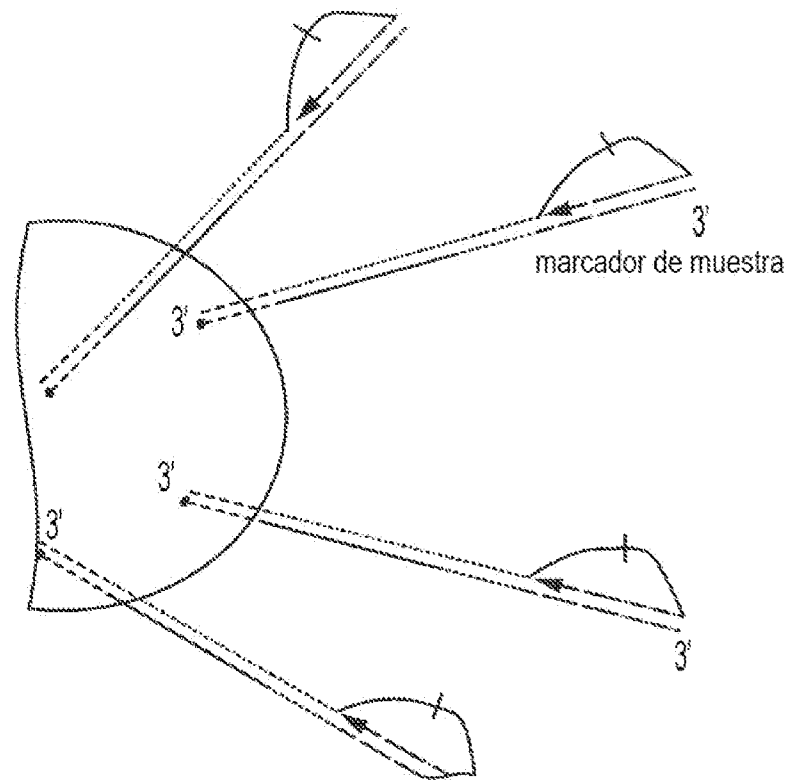


FIG. 14

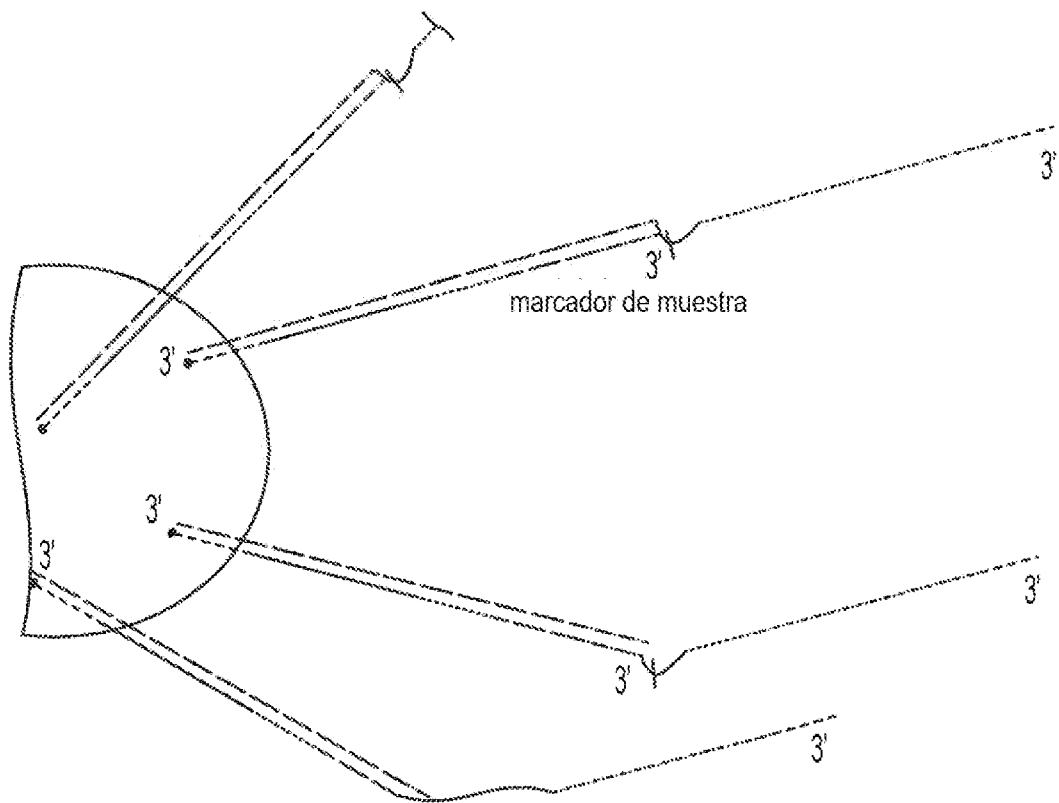


FIG. 15

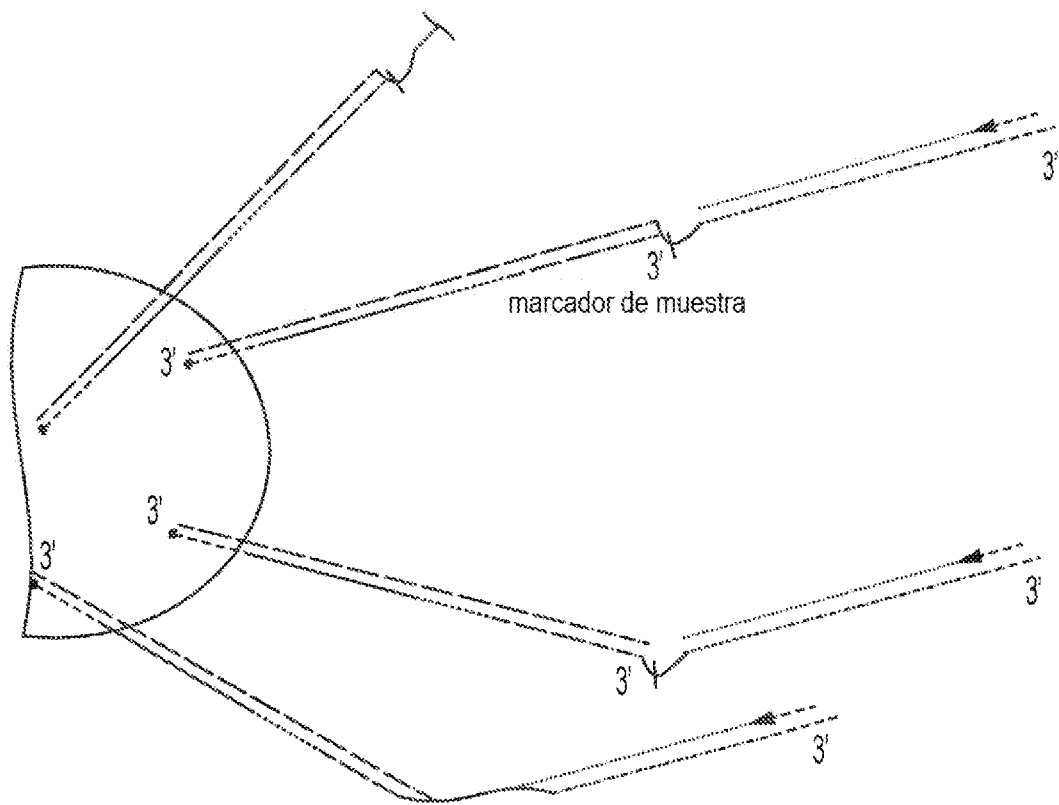


FIG. 16