

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 764 096**

51 Int. Cl.:

C40B 40/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.08.2014 PCT/US2014/051739**

87 Fecha y número de publicación internacional: **26.02.2015 WO15026853**

96 Fecha de presentación y número de la solicitud europea: **19.08.2014 E 14837282 (4)**

97 Fecha y número de publicación de la concesión europea: **23.10.2019 EP 3036359**

54 Título: **Bibliotecas de secuenciación de próxima generación**

30 Prioridad:

19.08.2013 US 201361867224 P

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

02.06.2020

73 Titular/es:

**ABBOTT MOLECULAR INC. (100.0%)
1300 East Touhy Avenue
Des Plaines, IL 60018, US**

72 Inventor/es:

KIM, DAE HYUN

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 764 096 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

Bibliotecas de secuenciación de próxima generación

5 **CAMPO DE LA INVENCION**

[0001] En este documento se proporciona tecnología relacionada con la secuenciación de próxima generación y particularmente, pero no exclusivamente, con métodos y composiciones para preparar una biblioteca de secuenciación de próxima generación que comprende fragmentos de ADN superpuestos y usar la biblioteca para secuenciar uno o más ácidos nucleicos diana.

ANTECEDENTES

[0002] Las secuencias de ácido nucleico codifican la información necesaria para que los seres vivos funcionen y se reproduzcan. Por lo tanto, determinar tales secuencias es una herramienta útil en la investigación pura sobre cómo y dónde viven los organismos, así como en ciencias aplicadas como el desarrollo de fármacos. En medicina, las herramientas de secuenciación se utilizan para el diagnóstico y el desarrollo de tratamientos para una variedad de patologías, que incluyen cáncer, enfermedades infecciosas, enfermedades cardíacas, trastornos autoinmunes, esclerosis múltiple y obesidad. En la industria, la secuenciación se utiliza para diseñar procesos enzimáticos y organismos sintéticos mejorados. En biología, tales herramientas se utilizan para estudiar la salud de los ecosistemas, p. ej., y por lo tanto tienen una amplia gama de utilidad.

[0003] Uno de los objetivos de la industria de la secuenciación se ha desplazado a la búsqueda de un mayor rendimiento y/o tecnologías de secuenciación de ácido nucleico de menor coste, a veces referido como tecnologías de secuenciación de "próxima generación" (NGS). Al hacer que la secuenciación sea de mayor rendimiento y/o menos costosa, el objetivo es hacer que la tecnología sea más accesible para la secuenciación. Estos objetivos se pueden alcanzar mediante el uso de plataformas y métodos de secuenciación que brindan preparación de muestras para grandes cantidades de muestras de complejidad significativa, secuenciación de grandes cantidades de muestras complejas y/o un alto volumen de generación y análisis de información en un corto período de tiempo. Varios métodos, como, p. ej., secuenciación por síntesis, secuenciación por hibridación y secuenciación por ligadura, están evolucionando para enfrentar estos desafíos.

[0004] Muchas plataformas de secuenciación de próxima generación (NGS) están disponibles para el alto rendimiento, secuenciación masivamente paralela de ácidos nucleicos. Muchos de estos sistemas, como los sistemas HiSeq y MiSeq producidos por Illumina, utilizan un enfoque de secuenciación por síntesis (SBS), en donde se determina una secuencia de nucleótidos utilizando detección e identificación base por base. Usando este enfoque particular, identificar 1 base requiere 1 ciclo del proceso químico de SBS (que puede involucrar cuatro reacciones separadas por lavados).

[0005] En la actualidad, estas tecnologías proporcionan una longitud de lectura máxima alcanzable de ~ 250 bases, que se puede extender a ~ 400 (2 x 250 bases con solapamiento suficiente para el montaje) si dos lecturas de extremo emparejado de alta calidad se adquieren de la misma plantilla y se ensamblan. Cada ciclo de SBS tarda aproximadamente 4 minutos en completarse; por lo tanto, en un enfoque de extremo emparejado para adquirir < 400 bases de información de secuencia, los 500 ciclos de SBS necesarios para producir las dos lecturas de < 250 bases tardan aproximadamente 37 horas en completarse. Además, la mayoría del rendimiento y la calidad de las tecnologías de secuenciación cíclica disminuyen sustancialmente después de determinar ~ 100 bases, introduciendo un grado de incertidumbre asociado con lecturas de secuencias individuales más largas que ~ 100 bases y los conjuntos de secuencias más largas en los que se utilizan. Debido a estas limitaciones de calidad y tiempo de las plataformas NGS actuales, las demandas cada vez mayores de secuencias de nucleótidos largas y de alta calidad están saturando las capacidades de salida de la base instalada de aparatos de secuenciación. En consecuencia, se necesitan tecnologías que proporcionen secuencias de alta calidad de ~ 500 bases o más a partir de un tiempo de secuenciación mucho más corto de varias horas en lugar de varios días.

[0006] US 2008/242560 se refiere a métodos para generar una matriz de secuencias de ácidos nucleicos amplificados. El-Sagheer et al (2011) Proceedings of the National Academy of Sciences 108: 11338-11343 se refiere a un enlazador de ADN artificial biocompatible que es leído por las polimerasas de ADN y es funcional en *Escherichia coli*.

SUMARIO

[0007] La invención se define en las reivindicaciones adjuntas.

[0008] Algunos intentos de adquirir secuencias más largas por la tecnología NGS han aplicado el enfoque de montaje de múltiples lecturas cortas para producir una secuencia más larga. Por ejemplo, la tecnología Moleculo proporcionada por Illumina inicialmente aísla una única copia de un fragmento de ADN largo (~ 10 Kbp). Este largo fragmento de ADN se amplifica clonalmente y posteriormente se fragmenta en piezas más pequeñas de aproximadamente 300-800

bases. Finalmente, los adaptadores con códigos de barras se agregan a las piezas más pequeñas utilizando una transposasa para generar la biblioteca de secuenciación. Se utiliza un protocolo SBS estándar para adquirir < 300-500 bases de secuencia de la plantilla diana (2 x 150 bases o 2 x 250 bases) y, una vez que se generan las secuencias, los códigos de barras se utilizan para analizar y ensamblar las lecturas para proporcionar la secuencia del ADN original de ~ 10 Kbp. Otro método implica la creación de una biblioteca de fragmentos superpuestos adecuada para un secuenciador Illumina, que produce lecturas que van desde ~ 400-460 bases al ensamblar dos lecturas de ~ 250-bases que se superponen en ~ 20-50 bases (ver, p. ej., Lundin, et al. (2012) Scientific Reports 3:1186). Esta biblioteca superpuesta se construye principalmente marcando fragmentos con secuencias de adaptador específicas, seguido de un paso de digestión y un proceso de selección de tamaño preciso. Por consiguiente, el documento en el presente documento es una tecnología para secuenciación que utiliza una longitud de lectura relativamente corta (p. ej., menos de 300 o menos de 200 bases, p. ej., ~ 30-50 bases) para lograr una secuencia contigua larga y de alta calidad comparable o superior a las tecnologías convencionales. A diferencia de las tecnologías convencionales, la tecnología proporcionada requiere solo un corto período de tiempo de ejecución (p. ej., ~ 3-4 horas) en un secuenciador (p. ej., la plataforma Illumina MiSeq), lo que disminuye drásticamente el tiempo dedicado al uso del aparato de secuenciación requerido para completar una secuencia de ejecución. Además, la tecnología da como resultado secuencias más largas (p. ej., ~ 500 bp a 1000 bp o más de secuencia de alta calidad) que la tecnología convencional. Además, el tiempo de ejecución no aumenta en función del tamaño del ácido nucleico que se va a secuenciar porque el tamaño de lectura corto (p. ej., ~ 30-50) permanece igual independientemente del tamaño del ácido nucleico que se va a secuenciar.

[0009] La tecnología no se limita a cualquier plataforma de secuenciación particular, pero es aplicable en general y la plataforma independiente. Por ejemplo, además de las disminuciones en el tiempo de ejecución en los sistemas Illumina, se logran reducciones de tiempo similares para las secuencias adquiridas utilizando, p. ej., los sistemas Life Technologies Ion Torrent y Qiagen GeneReader. En particular, mientras que la adquisición de una secuencia de base de ~ 400 utilizando la preparación de muestra convencional y la tecnología de secuenciación de Ion Torrent requiere aproximadamente 4 horas, la tecnología proporcionada en este documento reduce ese tiempo a aproximadamente 20 a 30 minutos. En algunas realizaciones, la tecnología es aplicable a los métodos basados en PCR en emulsión, basados en perlas y no basados, y por lo tanto encuentra uso en los sistemas SOLiD de Life Technologies y las plataformas de secuenciación Qiagen NGS.

[0010] Esta tecnología proporciona secuencia de alta calidad en una disminución del tiempo de secuenciación en relación con las tecnologías convencionales. La tecnología es independiente de la plataforma y, por lo tanto, es compatible con los aparatos de secuenciación existentes. La tecnología, en algunas realizaciones, mejora las plataformas NGS existentes al, p. ej., aumentar la longitud de lectura de las plataformas existentes y acortar el tiempo de adquisición de la secuencia. Además, una ventaja adicional de la tecnología actual es que reduce el consumo de reactivos de secuenciación caros y, por lo tanto, puede disminuir el costo total de secuenciación por base.

[0011] En resumen, la tecnología consiste en la producción de un conjunto de insertos de la biblioteca de secuencia corta superpuesta definidos (p. ej., menos de 300 o menos de 200 bases, p. ej., ~ 30-50 bases) de baldosas sobre una región de un ácido nucleico a ser secuenciado y compensado entre sí por, p. ej., 1-20, 1-10, o 1-5 bases (p. ej., en algunas realizaciones, por 1 base). Después de producir el conjunto de secuencias utilizando las bibliotecas superpuestas, los algoritmos de ensamblaje bioinformático se utilizan para "unir" el conjunto en mosaico de secuencias cortas superpuestas para producir la secuencia del ácido nucleico.

[0012] En primer lugar, la calidad de secuencia es alta porque cada base en el ácido nucleico a secuenciar se secuencia con alta cobertura (p. ej., 10 veces a 1000 veces la cobertura, p. ej., 50 veces a la cobertura 500 veces) dependiendo de la longitud de las secuencias cortas adquiridas y el desplazamiento entre secuencias en mosaico adyacentes. La alta frecuencia de muestreo en cada base minimiza o elimina los errores de secuencia al proporcionar una mayor información al proceso de ensamblaje que determina la identidad consensuada de cada base. Además, las primeras bases (p. ej., las primeras ~ 20-100 bases) determinadas en una secuencia generalmente tienen la mejor calidad. Por lo tanto, al usar estas bases iniciales determinadas durante la primera parte de cada ejecución de secuenciación (p. ej., las primeras ~ 30-50 bases), se utiliza información de secuencia de alta calidad en el ensamblaje. Por lo tanto, la tecnología minimiza los errores de secuencia, especialmente en aplicaciones donde se desean lecturas de secuencia largas que retienen la información de fase y enlace asociada con las lecturas y ensamblajes.

[0013] En segundo lugar, el tiempo del secuenciador se reduce porque determinar cada secuencia corta (p. ej., ~ 30-50 bases) requiere solo un pequeño número de ciclos de secuenciación (p. ej., 1 ciclo por base, p. ej., ~ 30-50 ciclos) en el aparato de secuencia. Al determinar todas las secuencias cortas en el conjunto de secuencias cortas en paralelo, el tiempo de secuenciación necesario para proporcionar la secuencia del ácido nucleico a secuenciar se reduce en gran medida, p. ej., de un octava a una décima parte del tiempo que necesitan las tecnologías convencionales para secuenciar el mismo ácido nucleico a secuenciar.

[0014] Esta tecnología para la preparación y la secuenciación de biblioteca NGS y el posterior análisis de corta lectura y montaje proporciona adquisición de más de ~ 500 bp (p. ej., 600, 700, 800 pares de bases o más) de secuencia contigua de alta calidad con la información de fase. La tecnología encuentra uso, p. ej., en la secuenciación de regiones

desconocidas a partir de una región conocida, p. ej., para interrogar variantes estructurales tales como translocaciones de genes, p. ej., la detección e identificación de pares de fusión de genes desconocidos. Además, la tecnología mejora las capacidades de secuenciación de las plataformas NGS existentes en relación con la longitud de lectura, el tiempo de ejecución y el costo sin actualizaciones y/o cambios en el hardware instalado existente y las químicas de secuenciación existentes.

[0015] En algunas realizaciones, la tecnología se relaciona con un método para determinar una secuencia de nucleótidos diana, el método que comprende la determinación de una primera subsecuencia de nucleótidos de la secuencia de nucleótidos diana, teniendo dicha subsecuencia de primer nucleótido que tiene un extremo 5' en x_1 de la secuencia de nucleótidos diana y que tiene un extremo 3' en el nucleótido y_1 de la secuencia de nucleótidos diana; determinar una segunda subsecuencia de nucleótidos de la secuencia de nucleótidos diana, dicha segunda subsecuencia de nucleótidos tiene un extremo 5' en el nucleótido x_2 de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_2 de la secuencia de nucleótidos diana; ensamblar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos para proporcionar una secuencia de consenso para la secuencia de nucleótidos diana, en donde $x_2 < y_1$; y $(y_1 - x_1) < 100$, $(y_2 - x_2) < 100$, y $(y_2 - y_1) < 5$. En algunas realizaciones, los fragmentos tienen menos de 100 bp, menos de 90 bp, menos de 80 bp, menos de 70 bp, menos de 60 bp, menos de 55 bp, menos de 50 bp, menos de 45 bp, menos de 40 bp o menos de 35 bp. Por consiguiente, en algunas realizaciones, $(y_1 - x_1) < 100$, 90, 80, 70, 60, 55, 50, 45, 40 o 35 y $(y_2 - x_2) < 100$, 90, 80, 70, 60, 55, 50, 45, 40 o 35. En algunas realizaciones, los fragmentos tienen menos de 50 bp; en consecuencia, en algunas realizaciones, $(y_1 - x_1) < 50$ e $(y_2 - x_2) < 50$.

[0016] En algunas realizaciones, el extremo 3' de los fragmentos difieren con respecto a la secuencia diana en menos de 4 o menos de 3 bases; en consecuencia, en algunas realizaciones, $(y_2 - y_1) < 4$ o $(y_2 - y_1) < 3$. En algunas realizaciones, los extremos 3' de los fragmentos difieren con respecto a la secuencia diana en 1 base; en consecuencia, en algunas realizaciones $(y_2 - y_1) = 1$.

[0017] En algunas realizaciones, un índice único (un "marcador" en algunas realizaciones) se utiliza para asociar un fragmento con el ácido nucleico de plantilla a partir del cual fue producido. En algunas realizaciones, un índice único es una secuencia única de nucleótidos sintéticos o una secuencia única de nucleótidos naturales que permite una fácil identificación del ácido nucleico diana dentro de una colección complicada de oligonucleótidos (p. ej., fragmentos) que contienen varias secuencias. En ciertas realizaciones, los identificadores de índice únicos se unen a fragmentos de ácido nucleico antes de unir las secuencias del adaptador. En algunas realizaciones, los identificadores de índice únicos están contenidos dentro de secuencias adaptadoras de modo que la secuencia única está contenida en las lecturas de secuenciación. Esto garantiza que se puedan detectar fragmentos homólogos en función de los índices únicos que se unen a cada fragmento, lo que proporciona una reconstrucción más inequívoca de una secuencia de consenso. Los fragmentos homólogos pueden aparecer, p. ej., por casualidad debido a repeticiones genómicas, dos fragmentos que se originan en cromosomas homólogos o fragmentos que se originan en ubicaciones superpuestas en el mismo cromosoma. Los fragmentos homólogos también pueden surgir de secuencias estrechamente relacionadas (p. ej., miembros de la familia de genes estrechamente relacionados, parálogos, ortólogos, ohnologos, xenólogos y/o pseudogenes). Dichos fragmentos pueden descartarse para garantizar que el ensamblaje de fragmentos largos se pueda calcular sin ambigüedades. Los marcadores pueden unirse como se describe anteriormente para las secuencias del adaptador. Los índices (p. ej., marcadores) pueden incluirse en las secuencias del adaptador.

[0018] En algunas realizaciones, el índice único (p. ej., identificador de índice, etiqueta, marcador, etc.) es un "código de barras". Como se usa en el presente documento, el término "código de barras" se refiere a una secuencia de ácido nucleico conocida que permite identificar alguna característica de un ácido nucleico con el que está asociado el código de barras. En algunas realizaciones, la característica del ácido nucleico a identificar es la muestra o fuente de la que se deriva el ácido nucleico. En algunas realizaciones, los códigos de barras tienen al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más nucleótidos de longitud. En algunas realizaciones, los códigos de barras son más cortos que 10, 9, 8, 7, 6, 5 o 4 nucleótidos de longitud. En algunas realizaciones, los códigos de barras asociados con algunos ácidos nucleicos tienen una longitud diferente que los códigos de barras asociados con otros ácidos nucleicos. En general, los códigos de barras tienen una longitud suficiente y comprenden secuencias que son lo suficientemente diferentes como para permitir la identificación de muestras basadas en códigos de barras con los que están asociados. En algunas realizaciones, un código de barras y la fuente de muestra con la que está asociado pueden identificarse con precisión después de la mutación, inserción o eliminación de uno o más nucleótidos en la secuencia del código de barras, como la mutación, inserción o eliminación de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más nucleótidos. En algunas realizaciones, cada código de barras en una pluralidad de códigos de barras difiere de cualquier otro código de barras en la pluralidad en dos o más posiciones de nucleótidos, como en 2, 3, 4, 5, 6, 7, 8, 9, 10 o más posiciones. En algunas realizaciones, uno o más adaptadores comprenden al menos una de una pluralidad de secuencias de códigos de barras. En algunas realizaciones, los métodos de la tecnología comprenden además identificar la muestra o fuente de la que se deriva un ácido nucleico diana en base a una secuencia de código de barras a la que se une el ácido nucleico diana. En algunas realizaciones, los métodos de la tecnología comprenden además identificar el ácido nucleico diana basado en una secuencia de código de barras a la que se une el ácido nucleico diana. Algunas realizaciones del método comprenden además identificar una fuente o muestra de la secuencia de nucleótidos diana determinando una secuencia de nucleótidos con código de barras. Algunas realizaciones del método comprenden además aplicaciones de conteo molecular (p. ej., enumeración de códigos de barras digitales y/o agrupamiento) para determinar los niveles

de expresión o el estado del número de copia de los objetivos deseados. En general, un código de barras puede comprender una secuencia de ácido nucleico que, cuando se une a un ácido nucleico diana, sirve como un identificador de la muestra de la que se deriva el polinucleótido diana.

5 **[0019]** En algunas realizaciones, los métodos proporcionan una secuencia de hasta 100 bases o, en algunas realizaciones, una secuencia de más de 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o más bases. En algunas realizaciones, la tecnología proporciona una secuencia de más de 1000 bases, p. ej., más de 2000, 2500, 3000, 3500, 4000, 4500 o 5000 o más bases. En algunas realizaciones, la secuencia consenso comprende hasta 100 bases o más, p. ej., 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más bases; en algunas realizaciones, la secuencia de consenso
10 comprende más de 1000 bases, p. ej., más de 2000, 2500, 3000, 3500, 4000, 4500 o 5000 o más bases.

[0020] La invención implica el uso de cebadores de oligonucleótidos, adaptadores, etc. que comprenden secuencias "universales". Una secuencia universal es una secuencia conocida, p. ej., para usar como un sitio de unión de cebador o sonda usando un cebador o sonda de una secuencia conocida (p. ej., complementaria a la secuencia universal).
15 Mientras que una secuencia específica de plantilla de un cebador, una secuencia de código de barras de un cebador y/o una secuencia de código de barras de un adaptador pueden diferir en las realizaciones de la tecnología, p. ej., de fragmento a fragmento, de muestra a muestra, de fuente a fuente, o de región de interés a región de interés, las realizaciones de la tecnología proporcionan que una secuencia universal sea la misma de fragmento a fragmento, de muestra a muestra, de fuente a fuente, o de región de interés a región de interés para que todos los fragmentos que
20 comprenden la secuencia universal pueden manejarse y/o tratarse de la misma manera o de manera similar, p. ej., amplificados, identificados, secuenciados, aislados, etc., usando métodos o técnicas similares (p. ej., usando el mismo cebador o sonda).

[0021] En realizaciones particulares, un cebador se utiliza que comprende una secuencia universal (p. ej., secuencia universal A), una secuencia de código de barras, y una secuencia de plantilla específica. Se usa un primer adaptador que comprende una secuencia universal B, y en realizaciones particulares, se usa un segundo adaptador que comprende una secuencia universal (p. ej., secuencia universal C). La secuencia universal A, la secuencia universal B y la secuencia universal C pueden ser cualquier secuencia. Esta nomenclatura es utilizada para observar que la secuencia universal A de un primer ácido nucleico (p. ej., un fragmento) que comprende la secuencia universal A es la misma que la secuencia universal A de un segundo ácido nucleico (p. ej., un fragmento) que comprende la secuencia universal A, la secuencia universal B de un primer ácido nucleico (p. ej., un fragmento) que comprende la secuencia universal B es la misma que la secuencia universal B de un segundo ácido nucleico (p. ej., un fragmento) que comprende la secuencia universal B, y la secuencia universal C de un primer ácido nucleico (p. ej., un fragmento) que comprende la secuencia universal C es la misma que la secuencia universal C de un segundo ácido nucleico (p. ej., un fragmento) que comprende secuencia universal C. Si bien las secuencias universales A, B y C son generalmente diferentes en las realizaciones de la tecnología, no es necesario que lo sean. Por lo tanto, en algunas realizaciones, las secuencias universales A y B son iguales; en algunas realizaciones, las secuencias universales B y C son iguales; en algunas realizaciones, las secuencias universales A y C son iguales; y en algunas realizaciones, las secuencias universales A, B y C son iguales. En algunas realizaciones, las secuencias universales A, B y C son diferentes.
25
30
35
40

[0022] Por ejemplo, si han de ser secuenciadas dos regiones de interés (p. ej., de la misma o de diferentes fuentes o, p. ej., a partir de dos regiones diferentes del mismo ácido nucleico, cromosoma, gen, etc.), dos cebadores pueden usados, un cebador que comprende una primera secuencia específica de plantilla para cebar desde la primera región de interés y un primer código de barras para asociar el primer producto amplificado con la primera región de interés y un segundo cebador que comprende una segunda secuencia específica de plantilla para cebar desde segunda región de interés y un segundo código de barras para asociar el segundo producto amplificado con la segunda región de interés. Sin embargo, estos dos cebadores, en algunas realizaciones, comprenderán la misma secuencia universal (p. ej., la secuencia universal A) para la agrupación y el procesamiento aguas abajo juntos. Se pueden usar dos o más secuencias universales y, en general, el número de secuencias universales será menor que el número de secuencias específicas de la diana y/o secuencias de código de barras para la agrupación de muestras y el tratamiento de agrupaciones como una sola muestra (lote).
45
50

[0023] En consecuencia, en la invención reivindicada, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende cebar a partir de una secuencia universal. La determinación de la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido 3'-O-bloqueado. Más específicamente, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido 3'-O-alquilílico, p. ej., en algunas realizaciones, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido 3'-O-propargilo. En algunas realizaciones, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido que comprende un terminador reversible.
55
60

[0024] Las lecturas de secuencia corta obtenidas se repartieron de acuerdo con su código de barras (p. ej., de-multiplexadas) y lecturas procedentes de las mismas muestras, las fuentes, las regiones de interés, etc. se han agrupado conjuntamente, p. ej., guardar en archivos separados o mantenidos en una estructura de datos organizada que permite identificar las lecturas agrupadas como tales. Luego, las secuencias cortas agrupadas se ensamblan en
65

una secuencia de consenso. El ensamblaje de secuencia generalmente se puede dividir en dos grandes categorías: ensamblaje de novo y ensamblaje de mapeo del genoma de referencia. En el ensamblaje de novo, las lecturas de secuencia se ensamblan juntas para que formen una secuencia nueva y previamente desconocida. En el mapeo del genoma de referencia, las lecturas de secuencia se ensamblan contra una secuencia principal existente (p. ej., una secuencia de referencia, etc.) para construir una secuencia que sea similar pero no necesariamente idéntica a la secuencia principal.

[0025] Por lo tanto, en algunas realizaciones, los ácidos nucleicos diana correspondientes a cada región de interés se reconstruyen utilizando un montaje de-novo. Para comenzar el proceso de reconstrucción, las lecturas cortas se unen bioinformáticamente al encontrar superposiciones y extenderlas para producir una secuencia de consenso. En algunas realizaciones, el método comprende además mapear la secuencia de consenso a una secuencia de referencia. Los métodos de la tecnología aprovechan la secuencia de puntajes de calidad que representan una base de llamadas de confianza para reconstruir fragmentos de longitud completa. Además del ensamblaje denovo, los fragmentos se pueden usar para obtener fases (asignación a copias homólogas de cromosomas) de variantes genómicas al observar que las secuencias de consenso se originan en cualquiera de los cromosomas.

[0026] En algunas realizaciones, un sistema de ordenador se implementa para el tratamiento de montaje y bioinformático de la secuencia de información (p. ej., la identificación de códigos de barras, particionar, organizar, hacer llamadas de base, la determinación de una identidad de consenso de cada base, lecturas de costura, la evaluación de las puntuaciones de calidad, lecturas de alineación y/o secuencias de consenso a una secuencia de referencia, etc.). En diversas realizaciones, un sistema informático incluye un bus u otro mecanismo de comunicación para comunicar información y un procesador acoplado con el bus para procesar información. En diversas realizaciones, el sistema informático incluye una memoria, que puede ser una memoria de acceso aleatorio (RAM) u otro dispositivo de almacenamiento dinámico, acoplado al bus, e instrucciones para que el procesador las ejecute. La memoria también se puede utilizar para almacenar variables temporales u otra información intermedia durante la ejecución de las instrucciones que debe ejecutar el procesador. En diversas realizaciones, el sistema informático incluye además una memoria de solo lectura (ROM) u otro dispositivo de almacenamiento estático acoplado al bus para almacenar información estática e instrucciones para el procesador. En algunas realizaciones, se proporciona un dispositivo de almacenamiento, tal como una unidad de estado sólido (p. ej., memoria "flash"), un disco magnético o un disco óptico, y se acopla al bus para almacenar información e instrucciones.

[0027] En diversas realizaciones, el sistema de ordenador está acoplado a través del bus a una pantalla, tal como un tubo de rayos catódicos (CRT) o de pantalla de cristal líquido (LCD), para mostrar información a un usuario de la computadora. En algunas realizaciones, un dispositivo de entrada, que incluye teclas alfanuméricas y otras, está acoplado al bus para comunicar información y selecciones de comandos al procesador. Otro tipo de dispositivo de entrada de usuario es un control de cursor, como un ratón, como una bola de seguimiento o teclas de dirección de cursor para comunicar información de dirección y selecciones de comandos al procesador y para controlar el movimiento del cursor en la pantalla.

[0028] En algunas realizaciones, un sistema de ordenador realiza aspectos de la presente tecnología. De acuerdo con ciertas realizaciones de la tecnología, el sistema informático proporciona resultados en respuesta al procesador que ejecuta una o más secuencias de una o más instrucciones contenidas en la memoria. Dichas instrucciones pueden leerse en la memoria desde otro medio legible por computadora, como el dispositivo de almacenamiento. Alternativamente, se pueden utilizar circuitos cableados en lugar de o en combinación con instrucciones de software para implementar la tecnología actual. Por lo tanto, las implementaciones de las presentes enseñanzas no se limitan a ninguna combinación específica de circuitos de hardware y software. Por ejemplo, como se describe en este documento, las realizaciones de la tecnología comprenden el uso de almacenamiento y transferencia de datos utilizando tecnología informática "en la nube", cableada (p. ej., fibra óptica, cable, cobre, ADSL, Ethernet y similares), y/o tecnología inalámbrica (p. ej., IEEE 802,11 y similares). Como se describe aquí, en algunas realizaciones, los componentes de la tecnología se conectan a través de una red de área local (LAN), una red de área local inalámbrica (WLAN), una red de área amplia (WAN) como Internet, o cualquier otro tipo de red, topología, y/o protocolo. En algunas realizaciones, la tecnología comprende el uso de un dispositivo portátil como una computadora de mano, un teléfono inteligente, una tableta, una computadora portátil, una computadora de mano, una computadora portátil, p. ej., para mostrar resultados, aceptar la entrada de un usuario, proporcionar instrucciones a otra computadora, almacenar datos y/o realizar otros pasos de los métodos proporcionados en este documento. Algunas realizaciones prevén el uso de un terminal de cliente ligero para mostrar resultados, aceptar entradas de un usuario, proporcionar instrucciones a otra computadora, almacenar datos y/o realizar otros pasos de los métodos proporcionados en este documento.

[0029] Algunas realizaciones proporcionan un método para determinar una secuencia de nucleótidos diana, comprendiendo el método la determinación de subsecuencias n de nucleótidos de la secuencia de nucleótidos diana (indexados sobre m), en donde el m -ésimo de nucleótidos subsecuencia tiene un extremo 5' en el nucleótido x_m de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_m de la secuencia de nucleótidos diana; la subsecuencia de nucleótidos $(m + 1)$ tiene un extremo 5' en el nucleótido $x_{m + 1}$ de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido $y_{m + 1}$ de la secuencia de nucleótidos diana; y ensamblar las subsecuencias de nucleótidos n para proporcionar una secuencia de consenso para la secuencia de nucleótidos diana, en donde m varía de 1 a n ; $x_{m + 1} < y_m$; y $(y_m - x_m) < 100, 90, 80, 70, 60, 50, 55, 50, 45, 40, 35$ o 30 o menos, $(y_{m + 1} - x_{m + 1}) < 100, 90, 80,$

5 70, 60, 50, 55, 50, 45, 40, 35 o 30 o menos, y $(y_{m+1} - y_m) < 20$, 10 o menos, o menos de 5, 4, o 3, o es igual a 1. En algunas realizaciones, los fragmentos tienen menos de 50 bp; en consecuencia, en algunas realizaciones $(y_m - x_m) < 50$ e $(y_{m+1} - x_{m+1}) < 50$. En algunas realizaciones, los fragmentos tienen menos de 40 bp; en consecuencia en algunas realizaciones $(y_m - x_m) < 40$ e $(y_{m+1} - x_{m+1}) < 40$. En algunas realizaciones, los fragmentos son menores de 30 bp; en consecuencia, en algunas realizaciones $(y_m - x_m) < 30$ e $(y_{m+1} - x_{m+1}) < 30$.

10 **[0030]** En algunas realizaciones, los extremos 3' de los fragmentos difieren en 4 o 3 bases con respecto a la secuencia de ácido nucleico diana. Por consiguiente, en algunas realizaciones $(y_{m+1} - y_m) < 4$ o $(y_{m+1} - y_m) < 3$. En algunas realizaciones, los extremos 3' de los fragmentos difieren en 1 base con respecto a la secuencia de ácido nucleico diana. Por lo tanto, en algunas realizaciones $(y_{m+1} - y_m) = 1$.

15 **[0031]** En algunas realizaciones, la determinación de las n subsecuencias de nucleótidos comprende cebado de una secuencia universal. En algunas realizaciones, determinar las subsecuencias de nucleótidos n comprende terminar la polimerización con un análogo de nucleótido 3'-O-bloqueado. En algunas realizaciones, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido de 3'-O-alquínilo. En algunas realizaciones, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido de 3'-O-propargilo. En algunas realizaciones, determinar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos comprende terminar la polimerización con un análogo de nucleótido que comprende un terminador reversible.

20 **[0032]** La invención reivindicada implica métodos para generar una biblioteca de secuenciación de próxima generación. En algunas realizaciones, los métodos comprenden amplificar una secuencia de nucleótidos diana usando un cebador que comprende una secuencia específica de diana, una secuencia universal A y una secuencia de nucleótidos con código de barras asociada con el ácido nucleico diana para proporcionar un amplicón identificable; ligar un primer oligonucleótido adaptador que comprende una secuencia universal B al extremo 3' del amplicón para formar un amplicón adaptador; circularizar el amplicón adaptador para formar una plantilla circular; generar una biblioteca de fragmentos de escalera a partir de la plantilla circular usando un análogo de nucleótido 3'-O-bloqueado; y ligar un segundo oligonucleótido adaptador que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar la biblioteca de secuenciación de próxima generación (p. ej., usando una ligasa o una ligadura química mediante, p. ej., química de clics, por ejemplo, una reacción catalizada por cobre de un alquino (p. ej., un 3' alquino) y una azida (p. ej., una 5' azida)).

25 **[0033]** En algunas realizaciones, la secuencia de nucleótidos de código de barras comprende de 1 a 20 nucleótidos. En algunas realizaciones, el primer oligonucleótido adaptador comprende de 10 a 80 nucleótidos. En algunas realizaciones, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera corresponden a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana y las secuencias de nucleótidos de los fragmentos tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana. En algunas realizaciones, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden menos de 100 nucleótidos, p. ej., menos de 90, 80, 70, 60, 50 o 40 nucleótidos, p. ej., 15 a 50, p. ej., 15 a 40 nucleótidos.

30 **[0034]** En algunas realizaciones, el primer oligonucleótido adaptador comprende un ADN de una sola cadena y/o el segundo oligonucleótido adaptador comprende un ADN monocatenario.

35 **[0035]** En algunas realizaciones generación de una biblioteca de fragmentos de escalera comprende la utilización de un oligonucleótido cebador complementario a la secuencia universal A.

40 **[0036]** En algunas realizaciones, los métodos comprenden además la amplificación de la biblioteca de secuenciación de próxima generación.

45 **[0037]** En algunas realizaciones, la 3'-O-alquínilo análogo de nucleótido es un análogo de nucleótido 3'-O-propargilo. En algunas realizaciones, el análogo de nucleótido comprende un terminador reversible.

50 **[0038]** La tecnología proporciona además métodos para determinar una secuencia de un ácido nucleico. Por ejemplo, en algunas realizaciones, el método comprende generar una biblioteca de secuenciación de próxima generación de acuerdo con la tecnología proporcionada en este documento; determinar una secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos de escalera, comprendiendo dicha secuencia de nucleótidos una subsecuencia de nucleótidos de la secuencia de nucleótidos diana; y determinar una secuencia de nucleótidos de código de barras del fragmento de la biblioteca de fragmentos de escalera.

55 **[0039]** En algunas realizaciones, la determinación de la secuencia de nucleótidos de un fragmento de la biblioteca comprende fragmentos de escalera usando un oligonucleótido cebador complementario a la secuencia universal de C. Además, en algunas realizaciones que determinan la secuencia de código de barras de nucleótidos del fragmento de la biblioteca comprende fragmentos de escalera usando un oligonucleótido cebador complementario a la secuencia universal B.

- 5 **[0040]** En algunas realizaciones la secuencia de nucleótidos de un fragmento de la biblioteca comprende fragmentos de escalera menos de 100 nucleótidos, p. ej., de 15 a 50 nucleótidos, p. ej., de 20 a 50, p. ej., de 25 a 50, p. ej., 30 a 50, p. ej., 35 a 50, p. ej., 40 a 50 nucleótidos. En algunas realizaciones, los métodos comprenden además asociar la secuencia de nucleótidos del código de barras con una fuente de la secuencia de nucleótidos diana.
- 10 **[0041]** En algunas realizaciones, los métodos comprenden además recoger secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera teniendo la misma secuencia de código de barras de nucleótidos. En algunas realizaciones, los métodos comprenden además ensamblar una pluralidad de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera para proporcionar una secuencia de consenso. En algunas realizaciones, los métodos comprenden además mapear la secuencia de consenso a una secuencia de referencia.
- 15 **[0042]** En algunas realizaciones, para permitir la reconstrucción de una secuencia consenso, la tecnología incluye unir etiquetas a los ácidos nucleicos, tales como proteínas de unión a ácidos nucleicos, etiquetas ópticas, análogos de nucleótidos, y otros conocidos en la técnica.
- 20 **[0043]** La tecnología proporciona composiciones relacionadas que comprenden una biblioteca de secuenciación de próxima generación, en donde la biblioteca de secuenciación de próxima generación comprende una pluralidad de ácidos nucleicos, comprendiendo cada ácido nucleico una secuencia universal A, una secuencia de código de barras de nucleótidos, una segunda secuencia universal B, una subsecuencia de nucleótidos de una secuencia de nucleótidos diana y una secuencia universal C. En algunas realizaciones, las composiciones comprenden n ácidos nucleicos, en donde la subsecuencia de nucleótidos enésima tiene un extremo 5' en el nucleótido X_m de la secuencia de nucleótidos diana y tiene un 3' terminar en el nucleótido y_m de la secuencia de nucleótidos diana; la subsecuencia de nucleótidos $(m + 1)$ tiene un extremo 5' en el nucleótido X_{m+1} de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_{m+1} de la secuencia de nucleótidos diana; m varía de 1 a n; $X_m = X_{m+1}$; y $(y_{m+1} - y_m) < 20$, 10 o menos de 5, 4, 3 o 2. En algunas realizaciones, los extremos 3' de los fragmentos de la biblioteca de secuenciación están desplazados entre sí y secuencia de nucleótidos diana por 4 o 3 bases; en consecuencia, en algunas realizaciones $(y_{m+1} - y_m) < 4$ o $(y_{m+1} - y_m) < 3$. En algunas realizaciones, los extremos 3' de los fragmentos de la biblioteca de secuenciación se compensan entre sí y la secuencia de nucleótidos diana por 1 base; en consecuencia, en algunas realizaciones $(y_{m+1} - y_m) = 1$.
- 30 **[0044]** En algunas realizaciones, la secuencia universal de B comprende de 10 a 100 nucleótidos y/o el nucleótido de código de barras secuencia comprende de 1 a 20 nucleótidos.
- 35 **[0045]** En algunas realizaciones, las composiciones comprenden además un análogo de nucleótido 3'-O-bloqueado, tales como un análogo de nucleótido de extremo 3'-O-alquínilo, p. ej., un análogo de nucleótido 3'-O-propargilo. En algunas realizaciones, las composiciones comprenden además un cebador de secuenciación. Por ejemplo, en algunas realizaciones, las composiciones comprenden además una secuenciación del cebador complementario a la secuencia universal de C y/o una secuenciación del cebador complementario a la secuencia universal de B.
- 40 **[0046]** En algunas realizaciones, la secuencia de código de barras de nucleótidos se asocia con la secuencia de nucleótidos diana. En algunas realizaciones, la pluralidad de ácidos nucleicos comprende ácidos nucleicos que tienen diferentes secuencias de nucleótidos de código de barras y subsecuencias de nucleótidos diferentes de una secuencia de nucleótidos diana, en donde cada secuencia de nucleótidos de código de barras está asociada con la secuencia de nucleótidos diana. En algunas realizaciones, la secuencia de nucleótidos del código de barras está asociada con la correspondencia uno a uno con la secuencia de nucleótidos diana.
- 45 **[0047]** En algunas realizaciones cada ácido nucleico de la biblioteca de secuenciación de próxima generación comprende un análogo de nucleótido de extremo 3'-O-bloqueado, p. ej., un análogo de nucleótido 3'-O-alquínilo, p. ej., un análogo de nucleótido 3'-O-propargilo. En algunas realizaciones, cada ácido nucleico de la biblioteca de secuenciación de próxima generación comprende un análogo de nucleótido que comprende un terminador reversible.
- 50 **[0048]** La tecnología proporcionada en el presente documento, en instancias de método y composición, encuentra uso, p. ej., para preparar una biblioteca NGS para la secuenciación, para adquirir una secuencia de nucleótidos, para asignar un polimorfismo de nucleótido único, para distinguir alelos, para secuenciar un genoma, para identificar variantes de poblaciones menores raras (p. ej., mutaciones somáticas en el cáncer o un patógeno contra un gran trasfondo de ADN del huésped o no patógeno), etc.
- 55 **[0049]** La secuenciación puede ser por cualquier método conocido en la técnica. En ciertas realizaciones, la secuenciación es secuenciación por síntesis. En otras realizaciones, la secuenciación es secuenciación de molécula única por síntesis. En ciertas realizaciones, la secuenciación implica hibridar un cebador con la plantilla para formar un dúplex plantilla/cebador, contactar el dúplex con una enzima polimerasa en presencia de nucleótidos marcados detectablemente en condiciones que permitan que la polimerasa agregue nucleótidos al cebador en una plantilla de manera dependiente, detectando una señal del nucleótido marcado incorporado, y repitiendo secuencialmente los pasos de contacto y detección al menos una vez, en donde la detección secuencial de nucleótidos marcados
- 60
- 65

incorporados determina la secuencia del ácido nucleico. Los ejemplos de marcadores detectables incluyen radiomarcadores, marcadores fluorescentes, marcadores enzimáticos, etc. En realizaciones particulares, el marcador detectable puede ser un marcador detectable ópticamente, tal como un marcador fluorescente. Los ejemplos de marcadores fluorescentes (para secuenciación y/u otros fines tales como el marcado de un ácido nucleico, cebador, sonda, etc.) incluyen cianina, rodamina, fluoresceína, cumarina, BODIPY, alexa o colorantes múltiples conjugados.

[0050] Algunas realizaciones proporcionan un método para generar una biblioteca de secuenciación de próxima generación, que comprende amplificar una secuencia de nucleótidos diana utilizando un cebador que comprende una secuencia específica diana, una secuencia universal de A, y una secuencia de código de barras de nucleótidos del método (p. ej., que comprende 1 a 20 nucleótidos) asociados con el ácido nucleico diana para proporcionar un amplicón identificable; ligar un primer oligonucleótido adaptador (p. ej., un ADN monocatenario, p. ej., que comprende de 10 a 80 nucleótidos) que comprende una secuencia universal B al extremo 3' del amplicón para formar un amplicón adaptador; circularizar el amplicón adaptador para formar una plantilla circular; generando a partir de la plantilla circular mediante el uso de un cebador complementario a la secuencia universal A y un análogo de nucleótido 3'-O-bloqueado (un análogo de nucleótido en 3'-O-alquínilo, p. ej., un análogo de nucleótido en 3'-O-propargilo, o que comprende un terminador reversible) una biblioteca de fragmentos de escalera que comprende una pluralidad de fragmentos; y ligar (p. ej., mediante química de clics, p. ej., usando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación, en donde las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden de 15 a 40 nucleótidos, las secuencias de nucleótidos de los fragmentos del fragmento de escalera la biblioteca corresponde a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana, y las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana.

[0051] Algunas realizaciones proporcionan un método para determinar una secuencia de nucleótidos diana, comprendiendo el método la amplificación de una secuencia de nucleótidos diana utilizando un cebador que comprende una secuencia específica diana, una secuencia universal de A, y una secuencia de código de barras de nucleótidos (p. ej., que comprenden de 1 a 20 nucleótidos) asociado con el ácido nucleico diana para proporcionar un amplicón; ligar un primer oligonucleótido adaptador (p. ej., un ADN monocatenario, p. ej., que comprende de 10 a 80 nucleótidos) que comprende una secuencia universal B al extremo 3' del amplicón para formar un amplicón adaptador; circularizar el amplicón adaptador para formar una plantilla circular; generando a partir de la plantilla circular mediante el uso de un cebador complementario a la secuencia universal A y un análogo de nucleótido 3'-O-bloqueado (un análogo de nucleótido en 3'-O-alquínilo, p. ej., un análogo de nucleótido en 3'-O-propargilo, o que comprende un terminador reversible) una biblioteca de fragmentos de escalera que comprende una pluralidad de fragmentos; ligar (p. ej., mediante química de clics, p. ej., utilizando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación; determinar una secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos de escalera (p. ej., usando un cebador oligonucleotídico complementario a la secuencia universal C), comprendiendo dicha secuencia de nucleótidos una subsecuencia de nucleótidos de la secuencia de nucleótidos diana; determinar una secuencia de nucleótidos de código de barras del fragmento de la biblioteca de fragmentos de escalera (p. ej., usando un cebador oligonucleotídico complementario a la secuencia universal B); asociar la secuencia de nucleótidos del código de barras con una fuente de la secuencia de nucleótidos diana; agrupamiento de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera que tienen la misma secuencia de nucleótidos de código de barras; ensamblar una pluralidad de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera para proporcionar una secuencia de consenso; y mapear la secuencia de consenso a una secuencia de referencia, en donde las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden 15 a 50, 15 a 40, o 15 a 30 nucleótidos, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera corresponden a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana, y la secuencia de consenso retiene información de fase y/o enlace del ácido nucleico diana.

[0052] La invención se refiere a métodos y composiciones para la secuenciación de un ácido nucleico (p. ej., por NGS) mediante la generación de una biblioteca de secuenciación de próxima generación utilizando nucleótidos modificados 3'-O-alquínilo. En algunas realizaciones, los nucleótidos modificados con 3'-O-alquínilo son nucleótidos 3'-O-propargilo (p. ej., 3'-O-propargilo-dNTP, p. ej., 3'-O-propargilo-dATP, 3'-O-propargilo -dCTP, 3'-O-propargilo-dGTP, 3'-O-propargilo-dTTP; véanse, p. ej., las solicitudes de patente de EE.UU. números de serie 14/463,412 y 14/463,416; y la aplicación de patente internacional. PCT/US2014/051726). Por ejemplo, las realizaciones de la tecnología están relacionadas con la generación de una biblioteca de secuenciación (p. ej., para NGS) que comprende una escalera de fragmentos de ácido nucleico producida mediante la incorporación de nucleótidos modificados con 3'-O-alquínilo que terminan la cadena por una polimerasa durante la síntesis in vitro de un ácido nucleico.

[0053] Las realizaciones particulares están relacionadas con la generación de una escalera de fragmento de ácido nucleico usando una reacción de polimerasa que comprende dNTPs estándar y 3'-O-propargilo-dNTPs en una relación molar de 1:500 a 500:1 (p. ej., una relación de dNTP estándar a 3'-O-propargilo-dNTP que es 1:500, 1:450, 1:400, 1:350, 1:300, 1:250, 1:200, 1:150, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 150:1, 200:1, 250:1, 300:1, 350:1, 400:1, 450:1 o 500:1). Los fragmentos de ácido nucleico terminados producidos por los métodos descritos en este documento comprenden un grupo propargilo en sus extremos 3'. Otras realizaciones están relacionadas con la fijación de un adaptador a los extremos 3' de los fragmentos de ácido nucleico usando conjugación química. Por ejemplo, en algunas realizaciones, un oligonucleótido modificado con 5'-azido (p. ej., un oligonucleótido modificado con 5'-azido-metilo) se conjuga con los fragmentos de ácido nucleico terminados en 3'-propargilo mediante química de clic (p. ej., en una reacción catalizada por un reactivo de cobre (p. ej., cobre (I)). En algunas realizaciones, una región diana se amplifica primero (p. ej., por PCR) para producir un amplicón diana para la secuenciación. En algunas realizaciones, la amplificación de la región diana comprende la amplificación de la región diana durante 5 a 15 ciclos (p. ej., una amplificación de "ciclo bajo").

[0054] Otras formas de realización prevén que el amplicón diana comprende una etiqueta (p. ej., comprende una secuencia de código de barras), p. ej., el amplicón diana es un amplicón identificable. En algunas realizaciones, un cebador utilizado en la amplificación de la región diana comprende una etiqueta (p. ej., que comprende una secuencia de código de barras) que se incorpora posteriormente al amplicón diana (p. ej., en una reacción de "copiar y etiquetar") para producir un amplicón identificable. En algunas realizaciones, un adaptador que comprende la etiqueta (p. ej., que comprende una secuencia de código de barras) se liga al amplicón diana después de la amplificación (p. ej., en una reacción de ligasa) para producir un adaptador-amplicón identificable. En algunas realizaciones, el cebador usado para producir un amplicón identificable en una reacción de copia y etiqueta comprende una región 3' que comprende una secuencia de cebado específica de la diana y una región 5' que comprende dos secuencias universales diferentes (p. ej., una secuencia universal A y una secuencia universal B) flanqueando una secuencia degenerada. En algunas realizaciones, un adaptador ligado a un amplicón para producir un adaptador-amplicón identificable es un adaptador bicatenario, p. ej., que comprende una cadena que comprende una secuencia degenerada (p. ej., que comprende 8 a 12 bases) flanqueada tanto en el extremo 5' como en el extremo 3' por dos secuencias universales diferentes (p. ej., una secuencia universal A y una secuencia universal B) y una segunda cadena que comprende una secuencia universal C (p. ej., en el extremo 5') y una secuencia (p. ej., en el extremo 3') que es complementario a la secuencia universal B y que tiene una T adicional en la posición 3'-terminal.

[0055] A continuación, las realizaciones de la tecnología proporcionan para la generación de fragmentos de escala de ácido nucleico del amplicón-adaptador, p. ej., para proporcionar una biblioteca de secuenciación para NGS. En particular, la tecnología proporciona la generación de una escalera de ácido nucleico terminado en 3'-O-propargilo-dN para la secuenciación de ácido nucleico (p. ej., NGS), p. ej., mediante el uso de una reacción de polimerasa que comprende dNTP estándar y 3'-O-propargilo-dNTP en una relación molar de 1:500 a 500:1 (dNTP estándar a 3'-O-propargilo-dNTP). Luego, en algunas realizaciones, la tecnología proporciona unir un adaptador a los extremos 3' de los fragmentos de ácido nucleico usando conjugación química. Por ejemplo, en algunas realizaciones, un oligonucleótido modificado con 5'-azido (p. ej., un oligonucleótido modificado con 5'-azido-metilo) se conjuga con los fragmentos de ácido nucleico terminados en 3'-propargilo mediante química de clic (p. ej., en una reacción catalizada por un reactivo de cobre (p. ej., cobre (I)).

[0056] Por consiguiente, algunas realizaciones proporcionan un método para generar una biblioteca de secuenciación de próxima generación, comprendiendo el método amplificar una secuencia de nucleótidos diana usando un cebador que comprende una secuencia específica diana, una secuencia universal A, una secuencia universal B y una secuencia de nucleótido de código de barras (p. ej., que comprende de 1 a 20 nucleótidos) asociada con el ácido nucleico diana para proporcionar un amplicón identificable; generar una escalera de fragmentos de ácido nucleico a partir del amplicón identificable usando un análogo de nucleótido 3'-O-bloqueado (p. ej., un análogo de nucleótido en 3'-O-alquínilo, un análogo de nucleótido en 3'-O-propargilo); y ligar (p. ej., mediante química de clics, p. ej., usando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación, en donde las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden de 15 a 100 nucleótidos, las secuencias de nucleótidos de los fragmentos del fragmento de escalera la biblioteca corresponde a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana, y las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana.

[0057] Algunos ejemplos dan a conocer un método para generar una biblioteca de secuenciación de próxima generación, comprendiendo el procedimiento la amplificación de una secuencia de nucleótidos diana para proporcionar un amplicón; ligar un adaptador (p. ej., un adaptador que comprende una hebra que comprende una secuencia degenerada (p. ej., que comprende 8 a 12 bases) flanqueada tanto en el extremo 5' como en el extremo 3' por dos secuencias universales diferentes (p. ej., una secuencia universal A y una secuencia universal B) y una segunda cadena que comprende una secuencia universal C (p. ej., en el extremo 5') y una secuencia (p. ej., en el extremo 3')

que es complementaria a la secuencia universal B y que tiene una T adicional en la posición 3'-terminal) al amplicón para producir un amplicón-adaptador; generar una escalera de fragmentos de ácido nucleico a partir del amplicón adaptador usando un análogo de nucleótido 3'-O-bloqueado (p. ej., un análogo de nucleótido en 3'-O-alquínilo, un análogo de nucleótido en 3'-O-propargilo); y ligar (p. ej., mediante química de clics, p. ej., usando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la escalera de biblioteca de fragmentos para generar una biblioteca de secuenciación de nueva generación, en donde las secuencias de nucleótidos de los fragmentos del fragmento de escalera biblioteca comprenden de 15 a 100 nucleótidos, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmento de escalera corresponde a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana, y las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana.

[0058] Algunas realizaciones proporcionan un método para determinar una secuencia de nucleótidos diana, que comprende el método de amplificación de una secuencia de nucleótidos diana utilizando un cebador que comprende una secuencia diana específica, una secuencia universal de A, una secuencia universal de B, y una secuencia de nucleótidos de código de barras (p. ej., que comprende de 1 a 20 nucleótidos) asociados con el ácido nucleico diana para proporcionar un amplicón identificable; generar una escalera de fragmentos de ácido nucleico a partir del amplicón identificable usando un análogo de nucleótido 3'-O-bloqueado (un análogo de nucleótido de 3'-O-alquínilo, p. ej., un análogo de nucleótido de 3'-O-propargilo); y ligar (p. ej., mediante química de clics, p. ej., usando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación; determinar una secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos de escalera (p. ej., usando un cebador oligonucleotídico complementario a la secuencia universal C), comprendiendo dicha secuencia de nucleótidos una subsecuencia de nucleótidos de la secuencia de nucleótidos diana; determinar una secuencia de nucleótidos de código de barras del fragmento de la biblioteca de fragmentos de escalera; asociar la secuencia de nucleótidos del código de barras con una fuente de la secuencia de nucleótidos diana; agrupamiento de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera que tienen la misma secuencia de nucleótidos de código de barras; ensamblar una pluralidad de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera para proporcionar una secuencia de consenso; y, en algunas realizaciones, mapear la secuencia de consenso a una secuencia de referencia, en donde las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden 15 a 50, 15 a 40, o 15 a 30 nucleótidos, las secuencias de nucleótidos de los fragmentos de escalera corresponden biblioteca de fragmentos de superposición de subsecuencias de nucleótidos dentro de la secuencia de nucleótidos diana, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana, y la secuencia de consenso conserva la eliminación y/o vinculación de información del ácido nucleico diana.

[0059] Algunos ejemplos dan a conocer un método para determinar una secuencia de nucleótidos diana, comprendiendo el método la amplificación de una secuencia de nucleótidos diana para proporcionar un amplicón; ligar un adaptador (p. ej., un adaptador que comprende una hebra que comprende una secuencia degenerada (p. ej., que comprende 8 a 12 bases) flanqueada tanto en el extremo 5' como en el extremo 3' por dos secuencias universales diferentes (p. ej., una secuencia universal A y una secuencia universal B) y una segunda cadena que comprende una secuencia universal C (p. ej., en el extremo 5') y una secuencia (p. ej., en el extremo 3') que es complementaria a la secuencia universal B y que tiene una T adicional en la posición de extremo 3') al amplicón para producir un adaptador-amplicón; generar una escalera de fragmentos de ácido nucleico a partir del amplicón adaptador usando un análogo de nucleótido 3'-O-bloqueado (p. ej., un análogo de nucleótido en 3'-O-alquínilo, un análogo de nucleótido en 3'-O-propargilo); y ligar (p. ej., mediante química de clics, p. ej., usando un reactivo catalítico a base de cobre, p. ej., para formar un triazol a partir de una azida y un alquínilo) un segundo oligonucleótido adaptador (p. ej., un ADN monocatenario) que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación; determinar una secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos de escalera (p. ej., usando un cebador oligonucleotídico complementario a la secuencia universal C), comprendiendo dicha secuencia de nucleótidos una subsecuencia de nucleótidos de la secuencia de nucleótidos diana; determinar una secuencia de nucleótidos de código de barras del fragmento de la biblioteca de fragmentos de escalera; asociar la secuencia de nucleótidos del código de barras con una fuente de la secuencia de nucleótidos diana; agrupamiento de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera que tienen la misma secuencia de nucleótidos de código de barras; ensamblar una pluralidad de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera para proporcionar una secuencia de consenso; y, en algunas realizaciones, mapear la secuencia de consenso a una secuencia de referencia, en donde las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera comprenden 15 a 50, 15 a 40, o 15 a 30 nucleótidos, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera corresponde a subsecuencias de nucleótidos superpuestas dentro de la secuencia de nucleótidos diana, las secuencias de nucleótidos de los fragmentos de la biblioteca de fragmentos de escalera tienen extremos 3' correspondientes a diferentes nucleótidos de la secuencia de nucleótidos diana, y la secuencia de consenso retiene información de fase y/o enlace del ácido nucleico diana.

[0060] Algunas realizaciones proporcionan un método para determinar una secuencia de nucleótidos diana, comprendiendo el método la determinación de una primera subsecuencia de nucleótidos de la secuencia de nucleótidos diana (p. ej., por el cebado de una secuencia universal y, p. ej., terminación de la polimerización con un análogo de nucleótido 3'-O-bloqueado tal como un análogo de nucleótido de 3'-O-alquinilo o un análogo de nucleótido de 3'-O-propargilo o polimerización de terminación con un análogo de nucleótido que comprende un terminador reversible), dicha primera subsecuencia de nucleótido tiene un extremo 5' en el nucleótido x1 de la secuencia de nucleótidos diana y que tiene un extremo 3' en el nucleótido y1 de la secuencia de nucleótidos diana; determinar una segunda subsecuencia de nucleótidos de la secuencia de nucleótidos diana (p. ej., cebando a partir de una secuencia universal y, p. ej., terminando la polimerización con un análogo de nucleótido 3'-O-bloqueado tal como un análogo de nucleótido con 3'-O-alquinilo o un análogo de 3'-O-propargilo o polimerización de terminación con un análogo de nucleótido que comprende un terminador reversible), dicha segunda subsecuencia de nucleótido tiene un extremo 5' en el nucleótido x2 de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y2 de la secuencia de nucleótidos diana; ensamblar la primera subsecuencia de nucleótidos y la segunda subsecuencia de nucleótidos para proporcionar una secuencia de consenso (p. ej., que comprende 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más de 1000, p. ej., 2000, 2500, 3000, 3500, 4000, 4500 o 5000, o más de 5000 bases) para la secuencia de nucleótidos diana; identificar una fuente o muestra de la secuencia de nucleótidos diana decodificando una secuencia de nucleótidos de código de barras; mapear la secuencia consenso (p. ej., retener información de fase y/o enlace del ácido nucleico diana) a una secuencia de referencia, en donde $x_2 < y_1$; y $(y_1 - x_1) < 100$ (p. ej., $(y_1 - x_1) < 90, 80, 70, 60, 55, 50, 45, 40, 35$ o 30), $(y_2 - x_2) < 100$ (p. ej., $(y_1 - x_1) < 90, 80, 70, 60, 55, 50, 45, 40, 35$ o 30), y $(y_2 - y_1) < 20$ (p. ej., $(y_2 - y_1) < 10, (y_2 - y_1) < 5, (y_2 - y_1) < 4, (y_2 - y_1) < 3, (y_2 - y_1) < 2$, o $(y_2 - y_1) = 1$).

[0061] Algunas realizaciones proporcionan un método para la determinación de una secuencia de nucleótido diana, comprendiendo el método determinar n subsecuencias de nucleótidos de la secuencia de nucleótidos diana (p. ej., cebando a partir de una secuencia universal y, p. ej., terminando la polimerización con un análogo de nucleótido 3'-O-bloqueado tal como un análogo de nucleótido en 3'-O-alquinilo o un análogo de nucleótido 3'-O-propargilo o polimerización de terminación con un análogo de nucleótido que comprende un terminador reversible), en donde la subsecuencia de nucleótidos enésima tiene un extremo 5' en el nucleótido X_m de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_m de la secuencia de nucleótidos diana; y la subsecuencia de nucleótidos $(m + 1)$ tiene un extremo 5' en el nucleótido X_{m+1} de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_{m+1} de la secuencia de nucleótidos diana; ensamblar las subsecuencias de n nucleótidos para proporcionar una secuencia de consenso (p. ej., que comprende 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o más de 1000 bases, p. ej., 2000, 2500, 3000, 3500, 4000, 4500 o 5000 o más de 5000 bases) para la secuencia de nucleótidos diana; identificar una fuente o muestra de la secuencia de nucleótidos diana decodificando una secuencia de nucleótidos de código de barras; y mapear la secuencia de consenso con una secuencia de referencia, en donde: m varía de 1 a n; $X_{m+1} < y_m$; y $(y_m - X_m) < 100$ (p. ej., $(y_m - X_m) < 90, 80, 70, 60, 55, 50, 45, 40, 35$ o 30), $(y_{m+1} - X_{m+1}) < 100$ (p. ej., $(y_{m+1} - X_{m+1}) < 90, 80, 70, 60, 55, 50, 45, 40, 35$ o 30), y $(y_{m+1} - y_m) < 20$ (p. ej., $(y_{m+1} - y_m) < 10, (y_{m+1} - y_m) < 5, (y_{m+1} - y_m) < 4, (y_{m+1} - y_m) < 3$, o $(y_{m+1} - y_m) = 1$) y la secuencia de consenso retiene la información de fase y/o enlace del ácido nucleico diana.

[0062] Algunas realizaciones de la tecnología proporcionan una composición para su uso como una biblioteca de secuenciación de próxima generación para obtener una secuencia de un ácido nucleico diana, la composición que comprende un análogo de nucleótido de 3'-O-bloqueado, un análogo de nucleótido de 3'-O-alquinilo, un análogo de nucleótido de 3'-O-propargilo, o un análogo de nucleótido que comprende un terminador reversible; un cebador de secuenciación (p. ej., complementario a una secuencia universal C); un segundo cebador de secuenciación (p. ej., complementario a una secuencia universal B); y n ácidos nucleicos, cada uno de los cuales comprende un análogo de nucleótido de 3'-O-alquinilo, o un análogo de nucleótido de 3'-O-propargilo unido (p. ej., mediante un enlace triazol formado, p. ej., mediante química de clic, p. ej., mediante una reacción entre una azida y un alquilo catalizado por un catalizador a base de cobre) a un adaptador (p. ej., un oligonucleótido adaptador de secuenciación de próxima generación), o un análogo de nucleótido que comprende un terminador reversible, en donde cada ácido nucleico comprende una subsecuencia de nucleótidos del ácido nucleico diana, una secuencia universal B que comprende de 10 a 100 nucleótidos, una secuencia universal C que comprende de 10 a 100 nucleótidos, y/o una secuencia de nucleótidos con código de barras que comprende de 1 a 20 nucleótidos, en donde la enésima subsecuencia de nucleótidos tiene un extremo 5' en el nucleótido X_m de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_m de la secuencia de nucleótidos diana; la subsecuencia de nucleótidos $(m + 1)$ tiene un extremo 5' en el nucleótido X_{m+1} de la secuencia de nucleótidos diana y tiene un extremo 3' en el nucleótido y_{m+1} de la secuencia de nucleótidos diana; m varía de 1 a n; $X_m = X_{m+1}$; $(y_{m+1} - y_m) < 20$ (p. ej., $(y_{m+1} - y_m) < 15, (y_{m+1} - y_m) < 10, (y_{m+1} - y_m) < 5, (y_{m+1} - y_m) < 4, (y_{m+1} - y_m) < 3$, o $(y_{m+1} - y_m) = 1$); los n ácidos nucleicos comprenden ácidos nucleicos que tienen diferentes secuencias de nucleótidos de código de barras y diferentes subsecuencias de nucleótidos de una secuencia de nucleótidos diana, en donde cada secuencia de nucleótidos de código de barras está asociada (p. ej., con correspondencia uno a uno) con una secuencia de nucleótidos diana.

[0063] La invención reivindicada proporciona una composición que comprende la superposición de fragmentos de ADN para su uso como una biblioteca de secuenciación de próxima generación para obtener una secuencia de un ácido nucleico diana, la composición que comprende ácidos n nucleicos (p. ej., una biblioteca de fragmento de ácido nucleico), en donde cada de los n ácidos nucleicos comprende un análogo de nucleótido 3'-O-bloqueado (un análogo

de nucleótido en 3'-O-alquínilo tal como un análogo de nucleótido en 3'-O-propargilo). Cada ácido nucleico de los n ácidos nucleicos comprende una subsecuencia de nucleótidos de una secuencia de nucleótidos diana. En particular, las realizaciones proporcionan una composición que comprende n ácidos nucleicos, en donde cada uno de los n ácidos nucleicos está terminado por un análogo de nucleótido 3'-O-bloqueado (un análogo de nucleótido en 3'-O-alquínilo tal como un análogo de nucleótido 3'-O-propargilo). Otras realizaciones proporcionan una composición que comprende n ácidos nucleicos (p. ej., una biblioteca de fragmentos de ácido nucleico), en donde cada uno de los n ácidos nucleicos comprende un análogo de nucleótido bloqueado en 3' (un análogo de nucleótido en 3'-O-alquínilo tal como un análogo de nucleótido 3'-O-propargilo) y cada uno de los n ácidos nucleicos se conjuga (p. ej., se une) a un adaptador de oligonucleótidos mediante un enlace triazol (p. ej., un enlace formado a partir de una conjugación química de un grupo propargilo y un grupo azido, p. ej., por una reacción química de clic). Por ejemplo, algunas realizaciones proporcionan una composición que comprende n ácidos nucleicos (p. ej., una biblioteca de fragmentos de ácido nucleico), en donde cada uno de los n ácidos nucleicos comprende un análogo de nucleótido 3'-O-propargilo (p. ej., un 3'-O-propargilo-dA, 3'-O-propargilo-dC, 3'-O-propargilo-dG, y/o 3'-O-propargilo-dT) conjugado (p. ej., unido) a un adaptador de oligonucleótidos mediante un enlace triazol (p. ej., un enlace formado a partir de una conjugación química de un grupo propargilo y un grupo azido, p. ej., mediante una reacción química de clic).

[0064] En algunas realizaciones, la composición para uso como una biblioteca de secuenciación de próxima generación para obtener una secuencia de un ácido nucleico diana es producido por un método que comprende sintetizar un ácidos nucleicos (p. ej., un fragmento de ácido nucleico de la biblioteca) utilizando una mezcla de dNTP y uno o más análogos de nucleótidos 3'-O-bloqueados (uno o más análogos de nucleótidos de 3'-O-alquínilo tales como uno o más análogos de nucleótidos de 3'-O-propargilo), p. ej., en una relación molar de 1:500 a 500:1 (p. ej., 1:500, 1:450, 1:400, 1:350, 1:300, 1:250, 1:200, 1:150, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 150:1, 200:1, 250:1, 300:1, 350:1, 400:1, 450:1 o 500:1). En algunas realizaciones, la composición se produce usando una polimerasa obtenida de, derivada de, aislada de, clonada de, etc. una especie *Thermococcus* (p. ej., un organismo del linaje taxonómico Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; *Thermococcus*). En algunas realizaciones, la polimerasa se obtiene de, derivada, aislada, clonada, etc. de una especie *Thermococcus* 9° N-7. En algunas realizaciones, la polimerasa comprende sustituciones de aminoácidos que proporcionan una incorporación mejorada de sustratos modificados tales como didesoxinucleótidos, ribonucleótidos y aciconucleótidos modificados. En algunas realizaciones, la polimerasa comprende sustituciones de aminoácidos que proporcionan una incorporación mejorada de análogos de nucleótidos que comprenden grupos funcionales 3' modificados tales como los dNTP de 3'-O-propargilo descritos aquí. En algunas realizaciones, la secuencia de aminoácidos de la polimerasa comprende una o más sustituciones de aminoácidos con respecto a *Thermococcus* sp. Secuencia de aminoácidos de polimerasa de tipo silvestre 9° N-7, p. ej., una sustitución de alanina por el ácido aspártico en la posición de aminoácido 141 (D141A), una sustitución de alanina por el ácido glutámico en la posición de aminoácido 143 (E143A), un sustitución de valina por la tirosina en la posición de aminoácidos 409 (Y409V), y/o una sustitución de leucina por la alanina en la posición de aminoácidos 485 (A485L). En algunas realizaciones, la polimerasa se proporciona en un organismo huésped heterólogo tal como *Escherichia coli* que comprende un *Thermococcus* sp clonado. Gen 9-N-7 polimerasa, p. ej., que comprende una o más mutaciones (p. ej., D141A, E143A, Y409V y/o A485L). En algunas realizaciones, la polimerasa es un *Thermococcus* sp. 9° N-7 polimerasa vendida bajo el nombre comercial THERMINATOR (p. ej., THERMINATOR II) por New England Biolabs (Ipswich, Mass.).

[0065] En consecuencia, la tecnología se refiere a mezclas de reacción que comprenden un ácido nucleico diana, una mezcla de dNTP y uno o más análogos de nucleótidos bloqueados en 3'-O (p. ej., uno o más análogos de nucleótido 3'-O-alquínilo tal como uno o más análogos de nucleótidos de 3'-O-propargilo), p. ej., en una relación molar de 1:500 a 500:1 (p. ej., 1:500, 1:450, 1:400, 1:350, 1:300, 1:250, 1:200, 1:150, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 150:1, 200:1, 250:1, 300:1, 350:1, 400:1, 450:1 o 500:1), y una polimerasa para sintetizar un ácido nucleico usando los dNTP y uno o más análogos de nucleótidos bloqueados en 3'-O (p. ej., una polimerasa obtenida de, derivada, aislada, clonada, etc. de una especie *Thermococcus*). En algunas realizaciones, el ácido nucleico diana es un amplicón. En algunas realizaciones, el ácido nucleico diana comprende un código de barras. En algunas realizaciones, el ácido nucleico diana es un amplicón que comprende un código de barras. En algunas realizaciones, el ácido nucleico diana es un amplicón ligado a un adaptador que comprende un código de barras. Algunas realizaciones proporcionan mezclas de reacción que comprenden una pluralidad de ácidos nucleicos diana, cada ácido nucleico diana comprende un código de barras asociado con una característica identificable del ácido nucleico diana.

[0066] Algunas formas de realización proporcionan una composición de la mezcla de reacción que comprende una plantilla (p. ej., una plantilla circular, p. ej., que comprende una secuencia de nucleótidos universal y/o una secuencia de código de barras de nucleótidos) que comprende una subsecuencia de un ácido nucleico diana, una polimerasa, uno o más fragmentos de una biblioteca de fragmentos de escalera y un análogo de nucleótido 3'-O-bloqueado.

[0067] Algunas formas de realización proporcionan una composición de la mezcla de reacción que comprende una biblioteca de ácidos nucleicos, la biblioteca de ácidos nucleicos que comprenden la superposición de secuencias de nucleótidos cortas sobre un ácido nucleico diana (p. ej., la secuencias de nucleótidos cortas superpuestas cubre una región de ácido nucleico diana que comprende 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 600 bases,

700 bases, 800 bases, 900 bases, 1000 bases o más de 1000 bases, p. ej., 2000 bases, 2500 bases, 3000 bases, 3500 bases, 4000 bases, 4500 bases, 5000 bases o más de 5000 bases) y compensadas entre sí por 1-20, 1-10 o 1-5 bases (p. ej., 1 base) y cada ácido nucleico de la biblioteca que comprende menos de 100 bases, menos de 90 bases, menos de 80 bases, menos de 70 bases, menos de 60 bases, menos de 50 bases, menos de 45 bases, menos de 40 bases, menos de 35 bases o menos de 30 bases.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0068] Estas y otras características, aspectos y ventajas de la presente tecnología se entenderán mejor con respecto a los siguientes dibujos:

La Figura 1 es un esquema que representa una realización de la tecnología para la secuenciación de un ácido nucleico.

La Figura 2 es un esquema que representa una realización de la tecnología para producir una biblioteca para la secuenciación de próxima generación. La Figura 2A muestra una realización de la tecnología y la Figura 2B muestra otra realización de la tecnología. La Figura 2C muestra otra realización de la tecnología.

La Figura 3 es un esquema que representa una realización de la tecnología para secuenciar un ácido nucleico.

La Figura 4 es un esquema que representa una realización de la tecnología para secuenciar un ácido nucleico.

La Figura 5 muestra diagramas de flujo relacionados con realizaciones de la tecnología que encuentran uso en la secuenciación de un ácido nucleico. La Figura 5A es un diagrama de flujo que muestra una realización de la tecnología que comprende obtener datos de secuencia de una biblioteca NGS y extraer las subsecuencias superpuestas de la secuencia diana. La Figura 5B es un diagrama de flujo que muestra una realización de la tecnología para extraer datos de secuencia que comprende concatenar archivos de datos de secuencia, identificar y extraer la secuencia diana y alinear las secuencias diana para proporcionar una secuencia de consenso.

La Figura 6 muestra la cobertura prevista y experimental de una secuencia diana por las lecturas de secuencia corta producidas por las realizaciones de la tecnología. La Figura 6A muestra la alineación de secuencia de lecturas de 40 bp y el perfil de cobertura de secuencia correspondiente. También se muestran las secuencias de consenso y de referencia (una secuencia de 177 bp que comprende el exón 2 de KRAS humano y secuencias de intrón flanqueantes parciales). La Figura 6B muestra la alineación de secuencia de lectura corta predicha y el perfil de cobertura de secuencia correspondiente para una secuencia de referencia de plantilla teórica.

La Figura 7 muestra un esquema de una realización de la tecnología relacionada con un esquema de "copiar y etiquetar" usando la extensión de polimerasa de un cebador que comprende una secuencia de código de barras y secuencias universales.

La Figura 8 muestra un esquema para la detección experimental de productos de reacción de "copiar y etiquetar" y para evaluar la efectividad del bloqueador de extensión de polimerasa.

La Figura 9 muestra un esquema para una estrategia de código de barras molecular basada en la ligadura del adaptador de acuerdo con realizaciones particulares de la tecnología.

La Figura 10 muestra un esquema para la detección experimental de productos ligados adaptadores.

La Figura 11 muestra un esquema para la ligadura intramolecular (circularización) de ADN monocatenario como un paso en la generación de fragmentos de escalera de acuerdo con la tecnología proporcionada en este documento.

La Figura 12 muestra un esquema para la detección experimental de plantillas circulares relacionadas con realizaciones de la tecnología relacionadas con la generación de plantillas circulares para la generación de fragmentos de escalera.

[0069] Ha de entenderse que las Figuras no están necesariamente dibujadas a escala, ni los objetos en las Figuras necesariamente han sido dibujados a escala en relación entre sí. Las Figuras son representaciones que pretenden aportar claridad y comprensión a varias instancias de aparatos, sistemas y métodos descritos aquí. Siempre que sea posible, se utilizarán los mismos números de referencia en todos los dibujos para referirse a las mismas partes o partes similares. Además, debe apreciarse que los dibujos no pretenden limitar el alcance de las presentes enseñanzas de ninguna manera.

DESCRIPCIÓN DETALLADA

[0070] La tecnología se refiere en general a la obtención de una secuencia de nucleótidos, tal como una secuencia consenso o una secuencia de haplotipos. En algunas realizaciones proporcionadas en el presente documento, se proporciona tecnología para producir una biblioteca de fragmentos de ADN superpuestos cortos a partir de un fragmento de ADN diana más grande para secuenciar. Los fragmentos cortos de ADN superpuestos tienen un rango de longitudes tales que un fragmento difiere de otro fragmento en 1-5 bases, preferiblemente 1 base, en sus extremos 3' (p. ej., una escalera de fragmentos similar a la producida por los métodos de secuenciación convencionales de Sanger). En algunas realizaciones, los fragmentos cortos de ADN superpuestos se indexan para generar una biblioteca de secuenciación de próxima generación (NGS). La biblioteca encuentra uso en la realización de NGS iniciando reacciones de secuenciación desde los extremos 3' variables de los fragmentos de ADN. La adquisición de lecturas de secuencia ~ 30- base a ~ 50-base de los extremos 3' de los cortos fragmentos solapantes produce un conjunto de

5 baldosas de lecturas de secuencia ~ 30-base a ~ 50-base que abarca el ADN diana más grande a ser secuenciado y desplazados uno otro por 1-5 bases, preferiblemente compensado por 1 base. El ensamblaje de las lecturas de secuencia corta superpuestas de - 30-50 bp produce una lectura contigua larga que cubre una región más grande (< 800-1000 bp) del fragmento de ADN diana. Por lo tanto, cada lectura de secuencia resulta de las bases de mayor calidad producidas por NGS (p. ej., las primeras 20-100 bases) y cada base del conjunto es el consenso de 30-50 lecturas de secuencia independientes de alta calidad.

10 **[0071]** En la descripción de esta tecnología, los encabezados de sección usados en este documento son sólo para fines de organización y no deben interpretarse como limitantes de la materia diana descrita en cualquier forma.

15 **[0072]** En esta descripción detallada de las diversas formas de realización, para fines de explicación, numerosos detalles específicos se exponen para proporcionar una comprensión completa de las realizaciones descritas. Sin embargo, un experto en la materia apreciará que estas diversas realizaciones se pueden practicar con o sin estos detalles específicos. En otros casos, las estructuras y dispositivos se muestran en forma de diagrama de bloques. Además, un experto en la materia puede apreciar fácilmente que las secuencias específicas en las que se presentan y realizan los métodos son ilustrativas y se contempla que las secuencias pueden variarse y aún permanecer dentro del espíritu y alcance de las diversas realizaciones descritas en el presente documento.

20 **[0073]** A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comúnmente entendido por un experto ordinario en la técnica a la que las diversas realizaciones descritas en este documento pertenece. Cuando las definiciones de los términos en las referencias parecen diferir de las definiciones proporcionadas en las presentes enseñanzas, prevalecerá la definición proporcionada en las presentes enseñanzas.

25 **Definiciones**

[0074] Para facilitar la comprensión de la tecnología actual, un número de términos y frases se definen a continuación. Se establecen definiciones adicionales a lo largo de la descripción detallada.

30 **[0075]** A lo largo de la memoria descriptiva y reivindicaciones, los siguientes términos tienen los significados asociados explícitamente en el presente documento, a menos que el contexto dicte claramente lo contrario. La frase "en una realización" como se usa en el presente documento no se refiere necesariamente a la misma realización, aunque puede hacerlo. Además, la frase "en otra realización" tal como se usa en el presente documento no se refiere necesariamente a otra realización, aunque puede referirse a ella. Por lo tanto, como se describe a continuación, se pueden combinar fácilmente diversas realizaciones de la invención, sin apartarse del alcance o espíritu de la invención.

35 **[0076]** Además, tal como se utiliza aquí, el término "o" es un operador inclusivo "o" y es equivalente a la expresión "y/o" a menos que el contexto claramente dicte otra cosa. El término "basado en" no es exclusivo y permite basarse en factores adicionales no descritos, a menos que el contexto indique claramente lo contrario. Además, a lo largo de la especificación, el significado de "un", "una", "el" y "la" incluye referencias plurales. El significado de "en" incluye "sobre" y "encima de".

40 **[0077]** Como se usa en este documento, un "nucleótido" comprende una "base" (alternativamente, una "nucleobase" o "base nitrogenada"), un "azúcar" (en particular, un azúcar de cinco carbonos, p. ej., ribosa o 2-desoxirribosa) y un "resto fosfato" de uno o más grupos fosfato (p. ej., un monofosfato, un difosfato o un trifosfato que consiste en uno, dos o tres fosfatos unidos, respectivamente). Sin el resto fosfato, la nucleobase y el azúcar componen un "nucleósido". Por lo tanto, un nucleótido también puede denominarse un nucleósido monofosfato o un nucleósido difosfato o un nucleósido trifosfato, dependiendo del número de grupos fosfato unidos. El resto fosfato generalmente está unido al carbono 5 del azúcar, aunque algunos nucleótidos comprenden restos fosfato unidos al carbono 2 o al carbono 3 del azúcar. Los nucleótidos contienen una purina (en los nucleótidos adenina y guanina) o una base de pirimidina (en los nucleótidos citosina, timina y uracilo). Los ribonucleótidos son nucleótidos en los que el azúcar es ribosa. Los desoxirribonucleótidos son nucleótidos en los que el azúcar es desoxirribosa.

45 **[0078]** Como se usa en este documento, un "ácido nucleico" se refiere a cualquier molécula de ácido nucleico, incluyendo, sin limitación, ADN, ARN, y sus híbridos. Las bases de ácido nucleico que forman moléculas de ácido nucleico pueden ser las bases A, C, G, T y U, así como sus derivados. Los derivados de estas bases son bien conocidos en la técnica. Debe entenderse que el término incluye, como equivalentes, análogos de ADN o ARN hechos a partir de análogos de nucleótidos. El término tal como se usa en el presente documento también abarca ADNc, que es ADN complementario, o copiado, producido a partir de una plantilla de ARN, p. ej., por la acción de una transcriptasa inversa.

50 **[0079]** Como se usa en este documento, "datos de secuenciación de ácidos nucleicos", "información de secuencia de ácido nucleico", "secuencia de ácido nucleico", "secuencia genómica", "secuencia genética", "secuencia de fragmento" o "lectura de secuenciación de ácidos nucleicos" denota cualquier información o datos que indiquen el orden de las bases de nucleótidos (p. ej., adenina, guanina, citosina y timina/uracilo) en una molécula (p. ej., un genoma completo, un transcriptoma completo, un exoma, oligonucleótido, polinucleótido, fragmento, etc.) de ADN o ARN.

[0080] Se debe entender que las presentes enseñanzas contemplan información de la secuencia obtenida usando todas las variedades disponibles de técnicas, plataformas o tecnologías, incluyendo, pero no limitado a: electroforesis capilar, microensayos, los sistemas basados en la ligación, los sistemas basados en polimerasa, sistemas basados en hibridación, sistemas de identificación de nucleótidos directos o indirectos, sistemas de pirosecuenciación, sistemas de detección basados en iones o pH, sistemas basados en firmas electrónicas, etc.

[0081] La referencia a una base, un nucleótido u otra molécula puede estar en singular o plural. Es decir, "una base" puede referirse a una sola molécula de esa base o a una pluralidad de la base, p. ej., en una solución.

[0082] Un "polinucleótido", "ácido nucleico" o "oligonucleótido" se refiere a un polímero lineal de nucleósidos (incluyendo desoxirribonucleósidos, ribonucleósidos, o análogos de los mismos) unidos por enlaces internucleosídicos. Típicamente, un polinucleótido comprende al menos tres nucleósidos. Usualmente, los oligonucleótidos varían en tamaño desde unas pocas unidades monoméricas, p. ej., 3-4, hasta varios cientos de unidades monoméricas. Siempre que un polinucleótido como un oligonucleótido esté representado por una secuencia de letras, como "ATGCCTG", se entenderá que los nucleótidos están en el orden 5' a 3' de izquierda a derecha y que "A" denota desoxiadenosina, "C" denota desoxicitidina, "G" denota desoxiguanosina y "T" denota timidina, a menos que se indique lo contrario. Las letras A, C, G y T pueden usarse para referirse a las bases mismas, a nucleósidos o a nucleótidos que comprenden las bases, como es estándar en la técnica.

[0083] Como se usa en el presente documento, el término "ácido nucleico diana" o "secuencia de nucleótidos diana" se refiere a cualquier secuencia de nucleótidos (p. ej., ARN o ADN), cuya manipulación puede ser considerada deseable por cualquier razón por un experto ordinario en la técnica. En algunos contextos, "ácido nucleico diana" se refiere a una secuencia de nucleótidos cuya secuencia de nucleótidos se va a determinar o se desea determinar. En algunos contextos, el término "secuencia de nucleótidos diana" se refiere a una secuencia en donde se genera un cebador o sonda parcial o completamente complementario.

[0084] Como se usa en este documento, el término "región de interés" se refiere a un ácido nucleico que se analiza (p. ej., usando una de las composiciones, sistemas o métodos descritos aquí). En algunas realizaciones, la región de interés es una porción de un genoma o región de ADN genómico (p. ej., que comprende uno o cromosomas o uno o más genes). En algunas realizaciones, se analiza el ARNm expresado desde una región de interés.

[0085] Tal como se utiliza aquí, el término "corresponde a" o "correspondiente" se usa en referencia a un ácido nucleico contiguo o secuencia de nucleótidos (p. ej., una subsecuencia) que es complementaria a, y así "corresponde a", la totalidad o una porción de una secuencia de ácido nucleico diana.

[0086] Como se usa en este documento, la frase "una pluralidad clonal de ácidos nucleicos" se refiere a los productos de ácido nucleico que son copias completas o parciales de un ácido nucleico de plantilla a partir del cual se generaron. Estos productos son sustancial o completamente o esencialmente idénticos entre sí, y son copias complementarias de la cadena de ácido nucleico de plantilla a partir de la cual se sintetizan, suponiendo que la tasa de incorporación incorrecta de nucleótidos durante la síntesis de las moléculas de ácido nucleico clonal es del 0%.

[0087] Como se usa en este documento, el término "biblioteca" se refiere a una pluralidad de ácidos nucleicos, p. ej., una pluralidad de diferentes ácidos nucleicos.

[0088] Como se usa en este documento, un "subsecuencia" de una secuencia de nucleótidos se refiere a cualquier secuencia de nucleótidos contenida dentro de la secuencia de nucleótidos, incluyendo cualquier subsecuencia que tiene un tamaño de una sola base hasta una subsecuencia que es una base más corta que la secuencia de nucleótidos.

[0089] Como se usa en este documento, el término "secuencia consenso" se refiere a una secuencia que es común a, o de otro modo presente en la fracción más grande, de un grupo alineado de secuencias. La secuencia de consenso muestra el nucleótido más comúnmente encontrado en cada posición dentro de las secuencias de ácido nucleico del grupo de secuencias. Una secuencia de consenso a menudo se "ensambla" a partir de lecturas de secuencia más cortas.

[0090] Como se usa en este documento, "montaje" se refiere a la generación de información de la secuencia de nucleótidos a partir de secuencias más cortas, p. ej., lecturas de secuencia adquiridas experimentalmente. El ensamblaje de secuencia generalmente se puede dividir en dos grandes categorías: ensamblaje nuevo y ensamblaje de mapeo del genoma de referencia. En el ensamblaje de novo, las lecturas de secuencia se ensamblan juntas para que formen una secuencia nueva y previamente desconocida. En el "mapeo" del genoma de referencia, las lecturas de secuencia se ensamblan contra una "secuencia de referencia" para construir una secuencia que es similar a, pero no necesariamente idéntica a la secuencia de referencia.

[0091] La frase "ejecución de secuenciación" se refiere a cualquier etapa o parte de un experimento de secuenciación realizado para determinar cierta información relativa a al menos una biomolécula (molécula p. ej., ácido nucleico).

[0092] Como se usa aquí, la frase "dNTP" significa trifosfato de deoxinucleótido, en donde el nucleótido comprende una base de nucleótido, tal como A, T, C, G o U.

5 **[0093]** El término "monómero", como se usa en este documento significa cualquier compuesto que se puede incorporar en una cadena molecular en crecimiento por una polimerasa dada. Dichos monómeros incluyen, sin limitaciones, nucleótidos naturales (p. ej., ATP, GTP, TTP, UTP, CTP, dATP, dGTP, dTTP, dUTP, análogos sintéticos), precursores para cada nucleótido, nucleótidos de origen no natural y sus precursores o cualquiera otra molécula que se puede incorporar en una cadena de polímero en crecimiento por una polimerasa dada.

10 **[0094]** Como se usa en el presente documento, "complementario" se refiere generalmente a dúplex específico de nucleótidos para formar los pares de bases Watson-Crick canónicos, como se entenderá por los expertos en la técnica. Sin embargo, complementario también incluye el emparejamiento de bases de análogos de nucleótidos que son capaces de emparejar bases universales con nucleótidos A, T, G o C y ácidos nucleicos bloqueados que mejoran la estabilidad térmica de los dúplex. Un experto en la materia reconocerá que la rigurosidad de la hibridación es un determinante en el grado de coincidencia o falta de coincidencia en el dúplex formado por hibridación.

15 **[0095]** Una "polimerasa" es una enzima generalmente para la unión de nucleótidos de 3'-OH 5'-trifosfato, oligómeros, y sus análogos. Las polimerasas incluyen, pero no se limitan a, polimerasas de ADN dependientes de ADN, polimerasas de ARN dependientes de ADN, polimerasas de ADN dependientes de ARN, polimerasas de ARN dependientes de ARN, polimerasa de ADN T7, polimerasa de ADN T3, polimerasa de ADN T4, polimerasa de ARN T7, T3 polimerasa de ARN, polimerasa de ARN SP6, polimerasa de ADN 1, fragmento Klenow, polimerasa de ADN *Thermophilus aquaticus* (Taq), polimerasa de ADN *Thermus thermophilus* (Tth), polimerasa de ADN Vent (New England Biolabs), polimerasa de ADN Deep Vent (New England Biolabs), *Bacillus stearothermophilus* (Bst) polimerasa de ADN, polimerasa de ADN de fragmento grande, Stoffel Fragment, polimerasa de ADN 9°N_m, polimerasa de 9°N_m, *Pyrococcus furiosus* (Pfu) polimerasa de ADN, *Thermus filiformis* (Tfi) polimerasa de ADN, polimerasa RepliPHI Phi29, *Thermococcus litoralis* (Tli) polimerasa de ADN, polimerasa beta ADN eucariota, telomerasa, polimerasa Terminator (p. ej., THERMINATOR I, THERMINATOR II, etc.) (New England Biolabs), KOD HiFi. polimerasa de ADN (Novagen), polimerasa de ADN KOD1, Q-beta replicasa, transferasa terminal, transcriptasa inversa AMV, transcriptasa inversa M-MLV, transcriptasa inversa Phi6, transcriptasa inversa VIH-1, nuevas polimerasas descubiertas por bioprospección y/o evolución molecular, y polimerasas citadas en la solicitud de patente de EE.UU. pub. N° 2007/0048748 y en la patente de EE.UU. N°s 6,329,178; 6,602,695; y 6,395,524. Estas polimerasas incluyen isoformas mutantes de tipo silvestre y variantes genéticamente modificadas como las exo-polimerasas; polimerasas con actividad de exonucleasa de corrección 3'→5' minimizada, indetectable y/o disminuida, y otros mutantes, p. ej., que toleran los nucleótidos marcados y los incorporan en una cadena de ácido nucleico. En algunas realizaciones, la polimerasa está diseñada para su uso, p. ej., en PCR en tiempo real, PCR de alta fidelidad, secuenciación de ADN de próxima generación, PCR rápida, PCR de arranque en caliente, PCR de muestra cruda, PCR robusta y/o diagnóstico molecular. Tales enzimas están disponibles de muchos proveedores comerciales, p. ej., enzimas de Kapa, Finnzymes, Promega, Invitrogen, Life Technologies, Thermo Scientific, Qiagen, Roche, etc.

20 **[0096]** El término "cebador" se refiere a un oligonucleótido, tanto si se produce de forma natural como en un resumen de restricción purificado o producido sintéticamente, que es capaz de actuar como un punto de inicio de síntesis cuando se coloca en condiciones en las que se induce la síntesis de un producto de extensión de cebador que es complementario a una cadena de ácido nucleico (p. ej., en presencia de nucleótidos y un agente inductor como la polimerasa de ADN y a una temperatura y pH adecuados). El cebador es preferiblemente monocatenario para la máxima eficiencia en la amplificación, pero alternativamente puede ser bicatenario. Si es de doble hebra, la imprimación se trata primero para separar sus hebras antes de usarse para preparar productos de extensión. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluida la temperatura, la fuente del cebador y el uso del método.

25 **[0097]** Como se usa en este documento, un "adaptador" es un oligonucleótido que está vinculado o está diseñado para ser unido a un ácido nucleico a introducir el ácido nucleico en un flujo de trabajo de secuenciación. Un adaptador puede ser monocatenario o bicatenario (p. ej., un ADN bicatenario o un ADN monocatenario). Como se usa en el presente documento, el término "adaptador" se refiere al adaptador nucleico en un estado que no está unido a otro ácido nucleico y en un estado que está unido a un ácido nucleico.

30 **[0098]** Al menos una parte del adaptador comprende una secuencia conocida. Por ejemplo, algunas realizaciones de adaptadores comprenden una secuencia de unión de cebador para la amplificación del ácido nucleico y/o para la unión de un cebador de secuenciación. Algunos adaptadores comprenden una secuencia para la hibridación de una sonda de captura complementaria. Algunos adaptadores comprenden un resto químico u otro (p. ej., un resto de biotina) para la captura y/o inmovilización en un soporte sólido (p. ej., que comprende un resto de avidina). Algunas realizaciones de adaptadores comprenden un marcador, índice, código de barras, etiqueta u otra secuencia por la cual el adaptador y un ácido nucleico al que está unido son identificables.

35 **[0099]** Algunos adaptadores comprenden una secuencia universal. Una secuencia universal es una secuencia

compartida por una pluralidad de adaptadores que de otro modo pueden tener diferentes secuencias fuera de la secuencia universal. Por ejemplo, una secuencia universal proporciona un sitio de unión de cebador común para una colección de ácidos nucleicos de diferentes ácidos nucleicos diana, p. ej., que pueden comprender diferentes códigos de barras.

5 [0100] Algunas realizaciones de adaptadores comprenden una secuencia definida pero desconocida. Por ejemplo, algunas realizaciones de adaptadores comprenden una secuencia degenerada de un número definido de bases (p. ej., una secuencia degenerada de 1 a 20 bases). Dicha secuencia se define incluso si no se conoce cada secuencia individual; sin embargo, dicha secuencia puede servir como índice, código de barras, etiqueta, etc., marcando
10 fragmentos de ácido nucleico de, p. ej., el mismo ácido nucleico diana.

[0101] Algunos adaptadores comprenden un extremo romo y algunos adaptadores comprenden un extremo con un saliente de una o más bases.

15 [0102] En realizaciones particulares proporcionadas en el presente documento, un adaptador comprende un resto azido, p. ej., el adaptador comprende un resto azido (p. ej., un azido-metilo) en su extremo 5'. Por lo tanto, algunas realizaciones están relacionadas con adaptadores que son o que comprenden un oligonucleótido modificado con 5'-azido u un oligonucleótido modificado con 5'-azido-metilo.

20 [0103] Como se usa en el presente documento, un "sistema" denota un conjunto de componentes, reales o abstractos, que comprende un todo en donde cada componente interactúa o está relacionado con al menos otro componente dentro del todo.

25 [0104] Como se usa en el presente documento, "índice" generalmente significará una marca o característica distintiva o de identificación. Un ejemplo de un índice es una secuencia de nucleótidos corta utilizada como un "código de barras" para identificar un nucleótido más largo que comprende el código de barras y otra secuencia.

30 [0105] Como se usa en el presente documento, el término "fase" o "de fase" se refiere al contenido único de los dos cromosomas heredados de cada padre y/o separando la información de secuencia derivada de la madre y el padre presente en un ácido nucleico (p. ej., un cromosoma). Por ejemplo, la información de fase haplotipo describe qué nucleótidos (p. ej., un SNP), regiones, porciones o fragmentos se originaron en cada uno de los cromosomas parentales (o están asociados con una cuasiespecie viral menor específica).

35 [0106] Como se usa en el presente documento, una "escalera de Sanger", "escalera de ADN", "escalera de fragmentos" o "escalera" se refiere a una biblioteca de ácidos nucleicos (p. ej., ADN) que difieren en longitud en un pequeño número de bases, p. ej., de una a cinco bases y en algunas realizaciones preferidas por una base. En algunas realizaciones, los ácidos nucleicos en la escalera tienen extremos 5' que corresponden a la misma posición de nucleótidos (o caen dentro de un pequeño rango de posiciones de nucleótidos, p. ej., 1-10 posiciones de nucleótidos) en la plantilla a partir de la cual se hicieron y tienen diferentes extremos 3' que corresponden a un rango de posiciones de nucleótidos en la plantilla a partir de la cual se hicieron. Ver, p. ej., escaleras ejemplares y/o escaleras similares a las proporcionadas en este documento en Sanger & Coulson (1975) "Un método rápido para determinar secuencias en ADN mediante síntesis cebada con polimerasa de ADN" J Mol Biol 94 (3): 441-8; Sanger et al (1977) "Secuenciación de ADN con inhibidores de terminación de cadena" Proc Natl Acad Sci EE.UU. 74 (12): 5463-7.

45 **Descripción**

[0107] En algunas realizaciones, la tecnología proporcionada en este documento proporciona métodos y composiciones para crear fragmentos cortos de ADN superpuestos que se extienden sobre una región más grande de fragmento de ADN. En particular, los fragmentos de ADN cortos componen una población de fragmentos de ADN que tienen un rango de tamaños que aumentan de tamaño de un fragmento al siguiente fragmento más grande, p. ej., de 1 a 20 pares de bases, 1 a 10 pares de bases, o 1 a 5 pares de bases, preferiblemente por 1 par de bases (p. ej., como en el caso de los fragmentos generados por la secuenciación de Sanger). En algunas realizaciones, se agrega un ácido nucleico corto que tiene una secuencia universal a los extremos 3' de cada fragmento (p. ej., el extremo del fragmento donde se genera la escalera). Posteriormente, los fragmentos se secuencian usando un cebador de secuenciación complementario a la secuencia universal. Como tal, las secuencias generadas tienen un rango de 5' (primeras) bases correspondientes a bases distribuidas a lo largo de la longitud del ADN más grande desde la primera base unida a la secuencia universal hasta 500 bases o más. Preferiblemente, las secuencias generadas tienen un rango de 5' (primeras) bases correspondientes a cada base distribuida a lo largo de la longitud del ADN más grande. Con este método, se utilizan lecturas cortas de NGS (< 30 a < 50 bases) para ensamblar una lectura contigua larga que retiene la información de fase y/o enlace (ver, p. ej., la Figura 1).

60 **1. Métodos para producir bibliotecas de NGS**

[0108] Las formas de realización de la tecnología se representan por el esquema mostrado en la Figura 2. En primer lugar, en algunas realizaciones, un ácido nucleico diana se amplifica usando uno o más cebadores específicos de la diana (ver, p. ej., Figura 2A, paso i; Figura 2C, paso i). El ácido nucleico diana puede ser un ADN o un ARN, p. ej., un

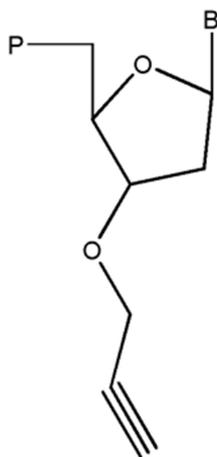
ADN genómico; ARNm; un cromosoma artificial cósmido, fosmídico o bacteriano (p. ej., que comprende un inserto), un gen, un plásmido, etc. En algunas realizaciones, primero se transcribe inversamente un ARN para producir un ADN. La amplificación puede ser PCR, ciclo limitado (ciclo bajo, p. ej., 5-15 ciclos (p. ej., 8 ciclos)) PCR, PCR isotérmica, amplificación con enzimas Phi29 o Bst, etc., p. ej., como se muestra en la Figura 2A y en la Figura 2C.

[0109] En algunas realizaciones, los cebadores específicos de la diana incluyen tanto una secuencia universal (p. ej., secuencia universal A) como una secuencia de índice de identificación única (p. ej., una secuencia de código de barras; ver Figura 2A, secuencia de código de barras "NNNNN") que permite el seguimiento y/o identificar el ácido nucleico diana a partir del cual se produjo el producto amplificado (amplicón). Generalmente, las secuencias de códigos de barras pueden constar de 1 a 10 o más nucleótidos. Por ejemplo, una secuencia de código de barras de 10 bases proporciona 1.048.576 (4^{10}) combinaciones de moléculas cebadoras específicas de objetivo identificables de forma única. En consecuencia, con una longitud de código de barras diseñada adecuadamente, un material de partida que contiene un número pequeño a muy grande de fragmentos de ADN diana puede etiquetarse e indexarse de manera confiable sin duplicar el etiquetado con la misma secuencia de código de barras.

[0110] En algunas realizaciones, los cebadores se usan para la amplificación (p. ej., no comprenden un código de barras) y el amplicón diana se liga a un adaptador que comprende una o más secuencias universales y/o una o más secuencias de códigos de barras (ver, p. ej., Figura 2C, secuencia de código de barras "NNNNNNNNNN", paso ii). Por lo tanto, en algunas realizaciones, el siguiente paso comprende ligar un adaptador al amplicón diana. En algunas realizaciones, el adaptador comprende la primera cadena que comprende un tramo de secuencia degenerada (p. ej., que comprende de 8 a 12 bases) flanqueada tanto en el extremo 5' como en el extremo 3' por dos secuencias universales diferentes (p. ej., secuencia universal A y secuencia universal B; ver Figura 9) y una segunda cadena que comprende una secuencia universal C (p. ej., en el extremo 5') y una secuencia (p. ej., en el extremo 3') que es complementaria a la secuencia universal B y que tiene una T adicional en la posición del terminal 3'.

[0111] Se proporcionan realizaciones en el presente documento para producir una escalera de fragmentos a partir de una plantilla circularizada (véase, p. ej., la Figura 2A y la Figura 2B) y se proporcionan realizaciones en el presente documento para producir una escalera de fragmentos a partir de una plantilla lineal (véase, p. ej., la Figura 2C). Por consiguiente, en algunas realizaciones, un siguiente paso comprende ligar los amplicones individuales con un código de barras único en sus extremos 3' a un oligonucleótido adaptador de aproximadamente 10 a 80 bases de longitud y que comprende una segunda secuencia universal (p. ej., secuencia universal B) (véase, p. ej., Figura 2A, paso ii). Después de la ligadura, los ácidos nucleicos de amplicón adaptador se autoligan (p. ej., circularizan) para formar una plantilla circular (véase, p. ej., la Figura 2A, paso iii). La circularización trae la secuencia universal en el extremo 3' adyacente a la secuencia del código de barras en el extremo 5'. La ligadura intramolecular puede realizarse usando una ligasa. Por ejemplo, CircLigase II (Epicentro) es una ligasa de ADN monocatenario termoestable que cataliza la ligadura intramolecular de plantillas de ADN monocatenario que tienen un grupo fosfato 5' y hidroxilo 3'.

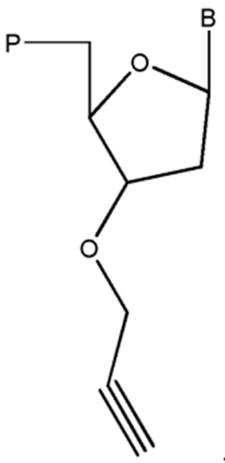
[0112] Luego, en realizaciones relacionadas con el uso de una plantilla circularizada, se genera una escalera de ADN similar a un fragmento Sanger mediante una reacción de polimerasa que usa un cebador complementario a la secuencia universal A y una mezcla de dNTP y análogos de dNTP bloqueados en 3'-O como descrito en este documento (véase, p. ej., la Figura 2A, paso iv). En algunas realizaciones, el análogo de dNTP 3'-O-bloqueado es un análogo de nucleótido 3'-O-alquínilo (p. ej., un grupo alquilo, que tiene una posición saturada (sp^3 hibridado) sobre un marco molecular junto a un grupo alquínilo, y variantes sustituidas de los mismos). En algunas realizaciones, el análogo de dNTP 3'-O-bloqueado es un análogo de nucleótido de 3'-O-propargilo que tiene una estructura como se muestra a continuación:



donde B es la base del nucleótido (p. ej., adenina, guanina, timina, citosina o una nucleobase natural o sintética, p. ej., una purina modificada tal como hipoxantina, xantina, 7-metilguanina; una pirimidina modificada tal como 5,6-dihidouracilo, 5-metilcitosina, 5-hidroximetilcitosina, etc.) y P comprende un resto fosfato. En algunas realizaciones,

P comprende un tetrafosfato; un trifosfato; un difosfato; un monofosfato; un 5'-hidroxilo; un alfa tiofosfato (p. ej., fosforotioato o fosforoditioato), un beta tiofosfato (p. ej., fosforotioato o fosforoditioato) y/o un gamma tiofosfato (p. ej., fosforotioato o fosforoditioato); o un alfa metilfosfonato, un beta metilfosfonato y/o un gamma metilfosfonato. La tecnología contempla otros grupos alquilino y encuentran uso en la tecnología, p. ej., butinilo, etc. En algunas realizaciones, el análogo de nucleótido es como se describe en otras secciones de la presente memoria.

[0113] Alternativamente, en realizaciones relacionadas con el uso de una plantilla lineal (véase, p. ej., la Figura 2C), se genera una escalera de ADN similar a un fragmento Sanger mediante una reacción de polimerasa usando un cebador complementario a una secuencia en el adaptador y una mezcla de dNTPs y análogos de dNTP 3'-O-bloqueados como se describe en el presente documento (véase, p. ej., Figura 2C, paso iii). El análogo dNTP3'-O-bloqueado es un análogo de nucleótido 3'-O-alquilino (p. ej., un alquilo que tiene una posición saturada (sp^3 hibridado) en un marco molecular junto a un grupo alquilino, y variantes sustituidas de los mismos). En algunas realizaciones, el análogo de dNTP 3'-O-bloqueado es un análogo de nucleótido de 3'-O-propargilo que tiene una estructura como se muestra a continuación:



donde B es la base del nucleótido (p. ej., adenina, guanina, timina, citosina o una nucleobase natural o sintética, p. ej., una purina modificada tal como hipoxantina, xantina, 7-metilguanina; una pirimidina modificada tal como 5,6-dihidouracilo, 5-metilcitosina, 5-hidroximetilcitosina, etc.) y P comprende un resto fosfato. En algunas realizaciones, P comprende un tetrafosfato; un trifosfato; un difosfato; un monofosfato; un 5' hidroxilo; un alfa tiofosfato (p. ej., fosforotioato o fosforoditioato), un beta tiofosfato (p. ej., fosforotioato o fosforoditioato) y/o un gamma tiofosfato (p. ej., fosforotioato o fosforoditioato); o un alfa metilfosfonato, un beta metilfosfonato y/o un gamma metilfosfonato. La tecnología contempla otros grupos alquilino y encuentran uso en la tecnología, p. ej., butinilo, etc. En algunas realizaciones, el análogo de nucleótido es como se describe en otras secciones de la presente memoria.

[0114] Las realizaciones de la tecnología proporcionan ventajas sobre las tecnologías existentes. Por ejemplo, en algunas realizaciones, la tecnología proporciona una secuencia de alta calidad a partir de una pequeña cantidad de ácido nucleico de entrada (p. ej., menos de 10 ng de ácido nucleico, p. ej., menos de 10 ng de ADN genómico). La tecnología proporciona el etiquetado robusto de plantillas individuales. La producción de bibliotecas es eficiente porque los métodos comprenden pocas manipulaciones (y por lo tanto pocos pasos de limpieza) y cada una de las manipulaciones tiene un rendimiento suficiente.

[0115] En algunos casos, el análogo de nucleótido comprende un terminador reversible que comprende un grupo de bloqueo que se puede eliminar para desbloquear el nucleótido. En algunas realizaciones, el análogo de nucleótido comprende un terminador funcional, p. ej., que proporciona una reactividad particular deseada para etapas posteriores.

[0116] Los análogos de nucleótidos dan como resultado la producción de una escalera fragmento que tiene fragmentos en un rango de tamaños. Por ejemplo, en algunas realizaciones, los fragmentos tienen longitudes que varían de aproximadamente 10 a aproximadamente 50 bp, aproximadamente 10 a aproximadamente 100 bp, y hasta aproximadamente 100 bp a aproximadamente 700 o aproximadamente 800 bp o más bp; además, en algunas realizaciones, se alcanzan longitudes mayores de 1000 bp ajustando la relación de dNTP y análogos de dNTP bloqueados en 3'O en la mezcla de reacción (p. ej., usando una relación de 1:500 a 500:1 (p. ej., 1:500, 1:450, 1:400, 1:350, 1:300, 1:250, 1:200, 1:150, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 150:1, 200:1, 250:1, 300:1, 350:1, 400:1, 450:1, o 500:1).

[0117] Las tecnologías de secuenciación de didesoxinucleótidos convencionales (ddNTP) (p. ej., las químicas de secuenciación de tipo Sanger) no son apropiadas para este paso en estas realizaciones debido a la falta de un grupo

3'-OH en la terminación ddNTP crea un extremo 3' terminal no reactivo que no puede aceptar la ligadura del segundo oligonucleótido adaptador en el paso siguiente.

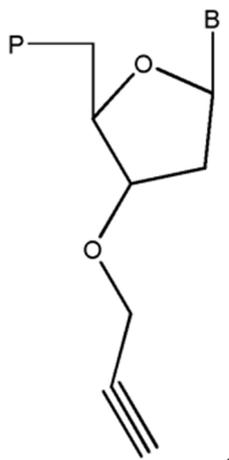
5 **[0118]** Una vez que la escalera del fragmento de ácido nucleico se genera con extremos 3' reactivos (p. ej., ligables), un segundo oligonucleótido adaptador que comprende una secuencia universal (p. ej., la secuencia universal C) se liga (enzimática o químicamente) a los extremos 3' de los fragmentos de la escalera de fragmentos de ácido nucleico para producir un NGS biblioteca. (ver, p. ej., Figura 2A, paso v; Figura 2C, paso (iv)). En algunas realizaciones, se realiza un ciclo limitado de PCR u otro método de amplificación para amplificar el producto final.

10 **[0119]** En algunas realizaciones, los métodos encuentran uso en la adquisición de secuencias cortas, p. ej., de < 120-200 bp. Dichas realizaciones encuentran uso, p. ej., para evaluar genes de cáncer, p. ej., para evaluar mutaciones de un panel de cáncer. En algunas realizaciones, la tecnología encuentra uso en la adquisición de secuencias de 500 bp, 1000 bp o más. Por ejemplo, en algunas realizaciones, un ácido nucleico diana se amplifica usando uno o más cebadores específicos diana (véase, p. ej., Figura 2B, paso i; Figura 2C, paso (i)). El ácido nucleico diana puede ser un ADN o un ARN, p. ej., un ADN genómico; ARNm; un cromosoma artificial cósmido, fosmídico o bacteriano (p. ej., que comprende un inserto), un gen, un plásmido, etc. En algunas realizaciones, primero se transcribe inversamente un ARN para producir un ADN. La amplificación puede ser PCR, PCR de ciclo limitado, PCR isotérmica, amplificación con enzimas Phi29 o Bst, etc., p. ej., como se muestra en la Figura 2B y en la Figura 2C.

20 **[0120]** En algunas realizaciones, los cebadores específicos del objetivo incluyen tanto una secuencia universal (p. ej., secuencia universal A) como una secuencia de índice de identificación única (p. ej., una secuencia de código de barras; ver Figura 2B, secuencia de código de barras "NNNNN") que permite el seguimiento y/o identificar el ácido nucleico diana a partir del cual se produjo el producto amplificado (amplicón). Generalmente, las secuencias de códigos de barras pueden constar de 1 a 10 o más nucleótidos. Por ejemplo, una secuencia de código de barras de 10 bases proporciona 1.048.576 (4^{10}) combinaciones de moléculas cebadoras específicas diana identificables de forma única. En consecuencia, con una longitud de código de barras diseñada adecuadamente, un material de partida que contiene un número pequeño a muy grande de fragmentos de ADN diana puede etiquetarse e indexarse de manera confiable sin duplicar el etiquetado con la misma secuencia de código de barras.

30 **[0121]** En algunas realizaciones, un siguiente paso comprende ligar los amplicones individuales con un código de barras único en sus extremos 3' a un oligonucleótido adaptador de aproximadamente 10 a 80 bases de longitud y que comprende una segunda secuencia universal (p. ej., secuencia universal B) (ver, p. ej., Figura 2B, paso ii). Después de la ligadura, los ácidos nucleicos del adaptador-amplicón se autoligan (p. ej., circularizan) para formar una plantilla circular (véase, p. ej., Figura 2B, paso iii). La circularización trae la secuencia universal en el extremo 3' adyacente a la secuencia del código de barras en el extremo 5'. La ligadura intramolecular puede realizarse usando una ligasa. Por ejemplo, CircLigase II (Epicentro) es una ligasa de ADN monocatenario termoestable que cataliza la ligadura intramolecular de plantillas de ADN monocatenario que tienen un fosfato 5' y un grupo hidroxilo 3'.

40 **[0122]** Usando la plantilla circularizada, se genera una escalera de ADN similar a un fragmento Sanger mediante una reacción de polimerasa que usa un cebador complementario a la secuencia universal A y una mezcla de dNTP y análogos de dNTP bloqueados en 3'-O como se describe en el presente documento (véase, p. ej., Figura 2B, paso iv). El análogo dNTP 3'-O-bloqueado es un análogo de nucleótido 3'-O-alquínilo (p. ej., un grupo alquilo, que tiene una posición saturada (sp^3 hibridado) sobre un marco molecular junto a un grupo alquínilo, y variantes sustituidas de los mismos). En algunas realizaciones, el análogo de dNTP 3'-O-bloqueado es un análogo de nucleótido de 3'-O-propargilo que tiene una estructura como se muestra a continuación:



65 donde B es la base del nucleótido (p. ej., adenina, guanina, timina, citosina o una nucleobase natural o sintética, p. ej., una purina modificada tal como hipoxantina, xantina, 7-metilguanina; una pirimidina modificada tal como 5,6-

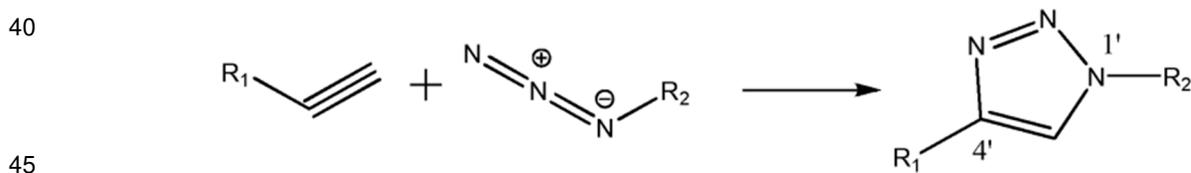
dihidouracilo, 5-metilcitosina, 5-hidroximetilcitosina, etc.) y P comprende un resto fosfato. En algunas realizaciones, P comprende un tetrafosfato; un trifosfato; un difosfato; un monofosfato; un 5' hidroxilo; un alfa tiofosfato (p. ej., fosforotioato o fosforoditioato), un beta tiofosfato (p. ej., fosforotioato o fosforoditioato), y/o un gamma tiofosfato (p. ej., fosforotioato o fosforoditioato); o un alfa metilfosfonato, un beta metilfosfonato y/o un gamma metilfosfonato. La tecnología contempla otros grupos alquínico y encuentran uso en la tecnología, p. ej., butinilo, etc. En algunas realizaciones, el análogo de nucleótido es como se describe en otras secciones de la presente memoria. La tecnología contempla otros grupos alquínico y encuentran uso en la tecnología, p. ej., butinilo, etc. En algunas realizaciones, el análogo de nucleótido es como se describe en otras secciones de la presente memoria.

[0123] En algunos casos, el análogo de nucleótido comprende un terminador reversible que comprende un grupo de bloqueo que se puede eliminar para desbloquear el nucleótido. En algunas realizaciones, el análogo de nucleótido comprende un terminador funcional, p. ej., que proporciona una reactividad particular deseada para etapas posteriores. Los análogos de nucleótidos dan como resultado la producción de una escalera de fragmentos que tiene fragmentos en un rango de tamaños. Por ejemplo, en algunas realizaciones, los fragmentos tienen longitudes que varían de < 100 bp a < 700 u 800 bp; además, en algunas realizaciones, se alcanzan longitudes de secuencia de más de 1000 bp a más de 10.000 bp, p. ej., ajustando la relación de dNTP y análogos de dNTP bloqueados en 3'-O en la mezcla de reacción.

[0124] Las tecnologías de secuenciación de didesoxinucleótidos convencionales (ddNTP) (p. ej., químicas de secuenciación de tipo Sanger) no son apropiadas para este paso en estas realizaciones porque la falta de un grupo 3'-OH en el ddNTP de terminación crea un terminal 3' no reactivo extremo que no puede aceptar la ligadura del segundo oligonucleótido adaptador en el siguiente paso.

[0125] A continuación, la escalera de fragmentos de ácido nucleico se circulariza para formar una biblioteca de círculos de ácido nucleico (véase, p. ej., la Figura 2B, paso v). Después de una digestión con una o más enzimas de restricción (véase, p. ej., la Figura 2B, paso vi), un segundo oligonucleótido adaptador (p. ej., que comprende una secuencia universal, p. ej., la secuencia universal C) se liga (enzimática o químicamente) al extremo 3' de los productos de digestión de la biblioteca de círculos de ácido nucleico para producir una biblioteca NGS. (ver, p. ej., Figura 2B, paso vii). En algunas realizaciones, se realiza un ciclo limitado de PCR u otro método de amplificación para amplificar el producto final. Sin limitarse a ningún método en particular o período de tiempo para realizar cualquiera de los pasos de los métodos proporcionados, en algunas realizaciones los métodos descritos toman de ~ 6 (p. ej., ~ 6,5) horas a ~ 9 (p. ej., ~ 8,5 horas) en completarse.

[0126] Los fragmentos comprenden un alquino 3'. Luego, en algunas realizaciones, el segundo oligonucleótido adaptador que comprende una secuencia universal (p. ej., secuencia universal de C) comprende un grupo azida 5' (N₃) que puede reaccionar con el grupo alquino fragmento 3'. Luego, en algunas realizaciones, se utiliza un proceso de "química de clic" tal como una cicloadición de azida-alquino para unir el adaptador al fragmento mediante la formación de un triazol:



donde R¹ y R² son individualmente cualquier estructura química o resto químico.

[0127] En algunas realizaciones, el enlace de anillo de triazol tiene una estructura de acuerdo con:

50

55

60

65

5

10

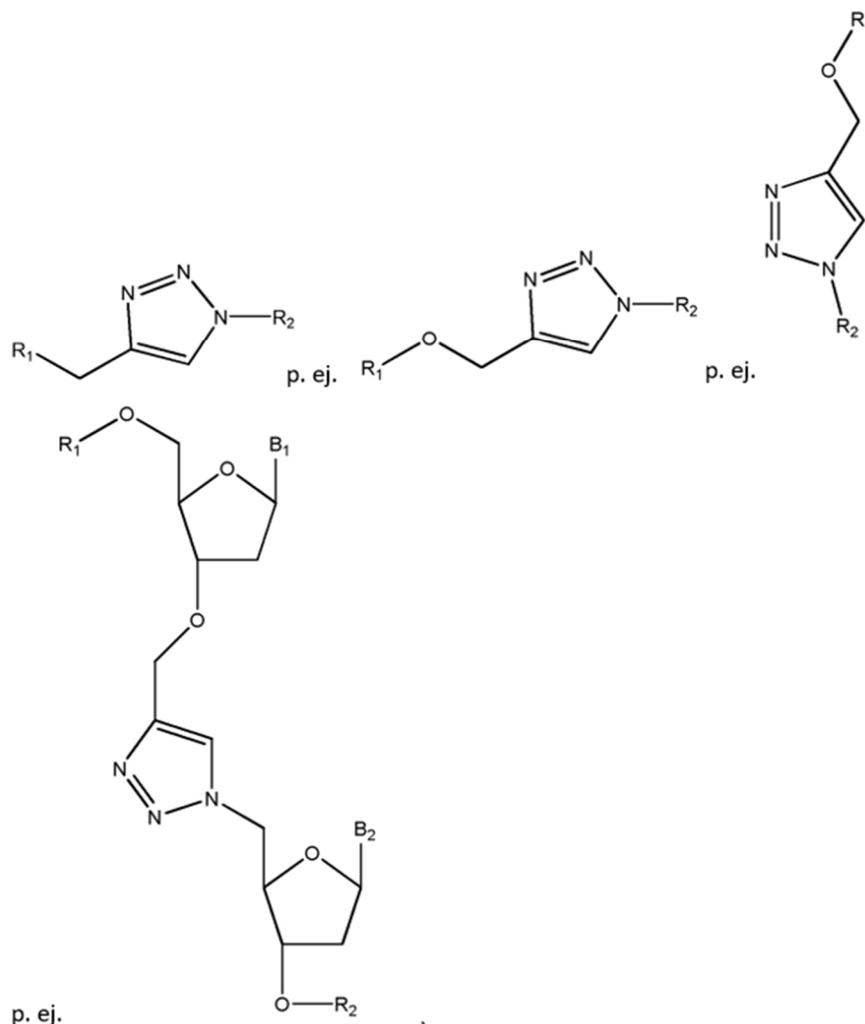
15

20

25

30

35



40 donde R^1 y R^2 son individualmente cualquier estructura química o resto químico (y no necesariamente el mismo de estructura a estructura) y B, B_1 y B_2 indican individualmente la base del nucleótido (p. ej., adenina, guanina, timina, citosina o una nucleobase natural o sintética, p. ej., una purina modificada como hipoxantina, xantina, 7-metilguanina; una pirimidina modificada como 5,6-dihidrouracilo, 5-metilcitosina, 5-hidroximetilcitosina, etc.).

45 **[0128]** El enlace de anillo de triazol formado por la cicloadición alquino-azida tiene características similares (p. ej., características físicas, químicas, biológicas) como un enlace fosfodiéster natural presente en los ácidos nucleicos y por lo tanto es un mímico de columna de ácido nucleico. En consecuencia, las enzimas convencionales que reconocen los ácidos nucleicos naturales como sustratos también reconocen como sustratos los productos formados por la cicloadición de alquino-azida según lo provisto por la tecnología aquí descrita. Ver, p. ej., El-Sagheer, et al. (2011)
50 "Enlace de ADN artificial biocompatible que es leído por las polimerasas de ADN y es funcional en *Escherichia coli*" Proc Natl Acad Sci EE.UU. 108 (28): 11338-43.

[0129] La biblioteca de fragmento NGS final se usa luego como entrada a un sistema NGS para la secuenciación. Durante la secuenciación, se secuencian ~ 20 a 50 bases de ADN adyacentes al adaptador que comprende la
55 secuencia universal C (correspondiente a ~ 20 a 50 bases del ácido nucleico diana) y se secuencia el código de barras adyacente al adaptador que comprende la secuencia universal B (ver, p. ej., Figura 3). Una vez que se obtienen las secuencias, las secuencias de códigos de barras analizan las lecturas de secuencia en contenedores para recopilar las lecturas de secuencia que se originaron a partir de una molécula plantilla etiquetada con esa secuencia de código de barras única en particular (véase, p. ej., la Figura 3). Las lecturas de secuencia en cada contenedor (para cada
60 secuencia de código de barras) se alinean entre sí y se ensamblan para construir una secuencia de consenso contigua más larga con la información de fase intacta. Esta secuencia se puede alinear con una secuencia de referencia apropiada para el análisis de secuencia aguas abajo.

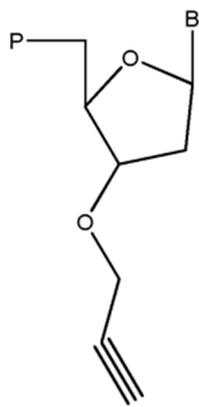
[0130] Se describen varios ejemplos de plataformas de secuenciación de ácido nucleico, ensamblaje de ácido nucleico y/o sistemas de mapeo de ácido nucleico (p. ej., software y/o hardware de computadora), p. ej., en la solicitud de
65 patente de EE.UU. pub. N° 2011/0270533. Las técnicas de "extremo emparejado", "pareja de pares" y otras secuencias

relacionadas con el ensamblaje son generalmente conocidas en la técnica de la biología molecular (Siegel AF et al., Genomics 2000, 68: 237-246; RoaCHJC et al., Genomics 1995, 26: 345-353). Estas técnicas de secuenciación permiten la determinación de múltiples "lecturas" de secuencia, cada una desde un lugar diferente en un único polinucleótido. Típicamente, se conoce la distancia entre las lecturas u otra información con respecto a una relación entre las lecturas. En algunas situaciones, estas técnicas de secuenciación proporcionan más información que la secuenciación de tramos múltiples de secuencias de ácido nucleico de manera aleatoria. Con el uso de herramientas de software apropiadas para el ensamblaje de información de secuencia (p. ej., Millikin SC et al., Genome Res. 2003, 13: 81-90; Kent, WJ et al., Genome Res. 2001, 11:1541-8) es posible hacer uso del conocimiento de que las secuencias no son completamente aleatorias, pero se sabe que ocurren a una distancia conocida y/o que tienen alguna otra relación, y por lo tanto están vinculadas en el genoma. Esta información puede ayudar en el ensamblaje de secuencias completas de ácido nucleico en una secuencia consenso.

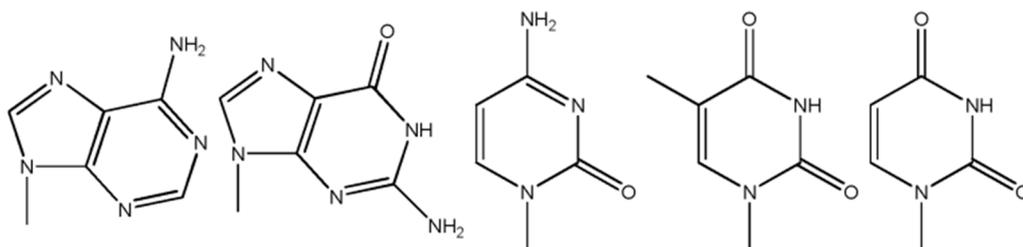
2. Análogos de nucleótidos

[0131] Un análogo de nucleótidos encuentra uso en la invención como un terminador de nucleótidos funcional (p. ej., en casos de composiciones, métodos, kits y sistemas descritos aquí). Un terminador de nucleótidos funcional termina la polimerización de un ácido nucleico, p. ej., bloqueando la participación del hidroxilo 3' en la reacción de polimerización, y comprende un grupo reactivo funcional que puede participar en otras reacciones químicas con otros grupos y restos químicos.

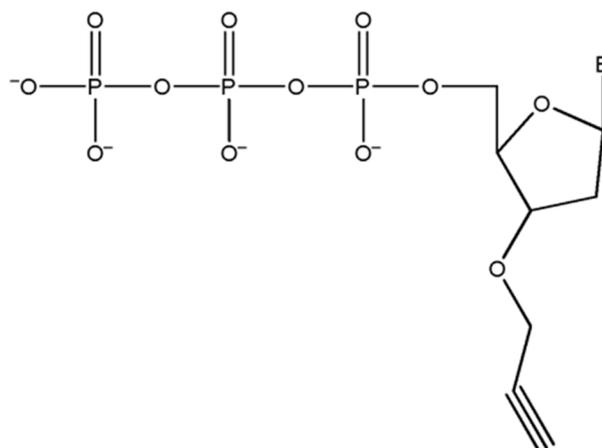
[0132] Un análogo de nucleótido que comprende un grupo alquínilo encuentra uso, p. ej., que tiene una estructura de acuerdo con:



en donde B es una base, p. ej., adenina, guanina, citosina, timina o uracilo, p. ej., que tiene una estructura de acuerdo con:



o una base o análogo modificado de una base, y P comprende un resto fosfato, p. ej., para proporcionar un nucleótido que tiene una estructura de acuerdo con:

5
10
15

[0133] En algunas realizaciones, P comprende un tetrafosfato; un trifosfato; un difosfato; un monofosfato; un 5' hidroxilo; un alfa tiofosfato (p. ej., fosforotioato o fosforoditioato), un beta tiofosfato (p. ej., fosforotioato o fosforoditioato) y/o un gamma tiofosfato (p. ej., fosforotioato o fosforoditioato); o un alfa metilfosfonato y/o un gamma metilfosfonato. En algunas realizaciones, P comprende una azida (p. ej., N₃, p. ej., N = N = N), proporcionando de este modo, en algunas realizaciones, un agente de polimerización bi-funcional direccional. En algunas realizaciones, la tecnología comprende el uso de un análogo de nucleótido como se describe en la solicitud de patente de EE.UU. N^{os} de Ser. 14/463,412 y 14/463,416; y solicitud de patente internacional PCT/US2014/051726.

[0134] El análogo de nucleótido es un análogo de nucleótido de 3'-O-alquino; en algunas realizaciones, el análogo de nucleótido es un análogo de nucleótido de 3'-O-propargilo tal como un 3'-O-propargilo-dNTP (en donde N = A, C, G, T o U). Un análogo de nucleótido de propargilo es un análogo de nucleótido que comprende una base (p. ej., adenina, guanina, citosina, timina o uracilo), una desoxirribosa y un resto químico alquino unido al oxígeno 3' de la desoxirribosa. La unión química entre los productos de extensión de polimerasa y los pares de conjugación apropiados (p. ej., moléculas modificadas con azida) se logra con alta eficiencia y especificidad utilizando, p. ej., química de clic.

[0135] El grupo 3' hidroxilo del análogo de nucleótido se tapa por un resto químico, p. ej., un alquino (p. ej., un triple enlace carbono-carbono), que detiene la elongación adicional del ácido nucleico (p. ej., ADN, ARN) de cadena cuando se incorpora por polimerasa (p. ej., ADN o polimerasa de ARN). El resto químico alquino es un par de conjugación de un grupo de azida bien conocido (N₃), p. ej., en una reacción de cicloadición de cobre 1,3-dipolar (I)-catalizada (p. ej., una reacción de "química clic"). La reacción del alquino con la azida forma un anillo de triazol de cinco miembros, lo que crea un enlace covalente. El enlace del anillo de triazol, en ciertas disposiciones posicionales, tiene características que son similares a un enlace fosfodiéster natural como se encuentra en una cadena principal de ácido nucleico convencional y, por lo tanto, el enlace triazol es una imitación de la cadena principal de ácido nucleico. Como se proporciona por algunas realizaciones en el presente documento, el uso de 3'-O-propargilo-dNTP crea fragmentos de ácido nucleico que tienen un grupo terminal 3'-O-alquino. Por consiguiente, estos fragmentos de ácido nucleico pueden ligarse químicamente usando química de clic a cualquier molécula modificada con azida, tal como oligonucleótidos modificados con 5'-azida (p. ej., adaptadores como se proporcionan aquí o un soporte sólido). El enlace químico triazol es compatible con las reacciones y enzimas típicas utilizadas para la bioquímica y la biología molecular y, como tal, no inhibe las reacciones enzimáticas. Por consiguiente, los fragmentos de ácido nucleico ligados químicamente pueden usarse luego en reacciones enzimáticas posteriores, tales como una reacción en cadena de la polimerasa, una reacción de secuenciación, etc.

[0136] En algunos casos, el análogo de nucleótido comprende un terminador reversible. Por ejemplo, en un análogo de nucleótido que comprende un terminador reversible, los grupos hidroxilo 3' están cubiertos con un resto químico que puede eliminarse con una reacción química específica, regenerando así un hidroxilo 3' libre. Como tal, algunos casos comprenden una reacción para eliminar el terminador reversible y, en algunas realizaciones, una etapa de purificación adicional para eliminar el resto de cubierta libre (terminador). En algunos casos, un nucleótido que comprende un terminador reversible es como se describe en la solicitud de patente de EE.UU. N^o de Serie 61/791,730 y/o en número de solicitud internacional PCT/US14/24391.

3. Adaptadores

[0137] Los métodos de la tecnología implican unir un adaptador a un ácido nucleico (p. ej., un amplicón o un fragmento de escalera como se describe aquí). En ciertas realizaciones, los adaptadores están unidos a un ácido nucleico con una enzima. La enzima puede ser una ligasa o una polimerasa. La ligasa puede ser cualquier enzima capaz de ligar un oligonucleótido (ARN monocatenario, ARN bicatenario, ADN monocatenario o ADN bicatenario) a otra molécula de ácido nucleico. Las ligasas adecuadas incluyen la ligasa de ADN T4 y la ligasa de ARN T4 (tales ligasas están disponibles comercialmente, p. ej., de New England BioLabs). Los métodos para usar ligasas son bien conocidos en

60
65

la técnica. La ligadura puede tener extremos romos o mediante el uso de extremos colgantes complementarios. En ciertas realizaciones, los extremos de los ácidos nucleicos pueden fosforilarse (p. ej., usando polinucleótido quinasa T4), repararse, recortarse (p. ej., usar una exonucleasa) o rellenarse (p. ej., usar una polimerasa y dNTP), para formar extremos romos. Al generar extremos romos, los extremos pueden tratarse con una polimerasa y dATP para formar una adición independiente de plantilla al extremo 3' de los fragmentos, produciendo así un solo saliente A. Esta A única se usa para guiar la ligadura de fragmentos con una sola T sobresaliendo del extremo 5' en un método denominado clonación TA. La polimerasa puede ser cualquier enzima capaz de agregar nucleótidos a los extremos 3' y 5' de las moléculas de ácido nucleico de plantilla.

[0138] En algunas realizaciones, un adaptador comprende un resto funcional para la unión química a un análogo de nucleótido. Por ejemplo, en algunas realizaciones, un adaptador comprende un grupo azida (p. ej., en el extremo 5') que es reactivo con un grupo alquínico (p. ej., un grupo propargilo, p. ej., en el extremo 3' de un ácido nucleico que comprende el análogo de nucleótido), p. ej., mediante una reacción química de clic (p. ej., usando un reactivo catalítico a base de cobre).

[0139] En algunas realizaciones, los adaptadores comprenden una secuencia universal y/o un índice, p. ej., una secuencia de nucleótidos de código de barras. Además, los adaptadores pueden contener uno o más de una variedad de elementos de secuencia, que incluyen, entre otros, una o más secuencias de recocido de cebadores de amplificación o sus complementos, una o más secuencias de recocido de cebadores de secuenciación o sus complementos, una o más secuencias de códigos de barras, una o más secuencias comunes compartidas entre múltiples adaptadores diferentes o subconjuntos de adaptadores diferentes (p. ej., una secuencia universal), uno o más sitios de reconocimiento de enzimas de restricción, uno o más salientes complementarios a uno o más salientes de polinucleótidos diana, uno o más sitios de unión de sonda (p. ej., para adjuntar a una plataforma de secuenciación, como una célula de flujo para secuenciación paralela masiva, como la desarrollada por Illumina, Inc.), una o más secuencias aleatorias o casi aleatorias (p. ej., uno o más nucleótidos seleccionados al azar de un conjunto de dos o más nucleótidos diferentes en una o más posiciones, con cada uno de los diferentes nucleótidos seleccionados en una o más posiciones representadas en un conjunto de adaptadores que comprende la secuencia aleatoria), y combinaciones de los mismos. Dos o más elementos de secuencia pueden ser no adyacentes entre sí (p. ej., separados por uno o más nucleótidos), adyacentes entre sí, parcialmente solapados o completamente solapados. Por ejemplo, una secuencia de recocido de cebador de amplificación también puede servir como secuencia de recocido de cebador de secuenciación. Los elementos de secuencia se pueden ubicar en o cerca del extremo 3', en o cerca del extremo 5', o en el interior del oligonucleótido adaptador. Cuando un oligonucleótido adaptador es capaz de formar una estructura secundaria, como una horquilla, los elementos de secuencia pueden ubicarse parcial o completamente fuera de la estructura secundaria, parcial o completamente dentro de la estructura secundaria, o entre secuencias que participan en la estructura secundaria. Por ejemplo, cuando un oligonucleótido adaptador comprende una estructura de horquilla, los elementos de secuencia pueden ubicarse parcial o completamente dentro o fuera de las secuencias hibridables (el "tallo"), incluso en la secuencia entre las secuencias hibridables (el "bucle"). En algunas realizaciones, los primeros oligonucleótidos adaptadores en una pluralidad de primeros oligonucleótidos adaptadores que tienen diferentes secuencias de códigos de barras comprenden un elemento de secuencia común entre todos los primeros oligonucleótidos adaptadores en la pluralidad. En algunas realizaciones, todos los segundos oligonucleótidos adaptadores comprenden un elemento de secuencia común entre todos los segundos oligonucleótidos adaptadores que es diferente del elemento de secuencia común compartido por los primeros oligonucleótidos adaptadores. Una diferencia en los elementos de la secuencia puede ser tal que al menos una parte de los diferentes adaptadores no se alineen completamente, p. ej., debido a cambios en la longitud de la secuencia, eliminación o inserción de uno o más nucleótidos, o un cambio en la composición de nucleótidos en una o más posiciones de nucleótidos (como un cambio de base o modificación de base). En algunas realizaciones, un oligonucleótido adaptador comprende un saliente 5', un saliente 3', o ambos, que es complementario a uno o más polinucleótidos diana. Los salientes complementarios pueden tener uno o más nucleótidos de longitud, incluidos, entre otros, pero sin limitarse a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más nucleótidos de longitud. Los salientes complementarios pueden comprender una secuencia fija. Los salientes complementarios pueden comprender una secuencia aleatoria de uno o más nucleótidos, de modo que uno o más nucleótidos se seleccionen al azar de un conjunto de dos o más nucleótidos diferentes en una o más posiciones, con cada uno de los diferentes nucleótidos seleccionados en una o más posiciones representada en un grupo de adaptadores con salientes complementarios que comprenden la secuencia aleatoria. En algunas realizaciones, un saliente del adaptador es complementario a un saliente del polinucleótido diana producido por digestión con endonucleasa de restricción. En algunas realizaciones, un saliente adaptador consiste en una adenina o una timina.

[0140] En algunas realizaciones, las secuencias adaptadoras pueden contener un elemento de identificación de sitio de unión molecular para facilitar la identificación y el aislamiento del ácido nucleico diana para aplicaciones posteriores. La unión molecular como mecanismo de afinidad permite que la interacción entre dos moléculas dé como resultado un complejo de asociación estable. Las moléculas que pueden participar en las reacciones de unión molecular incluyen proteínas, ácidos nucleicos, carbohidratos, lípidos y pequeñas moléculas orgánicas como ligandos, péptidos o medicamentos.

[0141] Cuando se usa un sitio de unión molecular de ácido nucleico como parte del adaptador, se puede usar para emplear hibridación selectiva para aislar una secuencia diana. La hibridación selectiva puede restringir la hibridación

sustancial a los ácidos nucleicos diana que contienen el adaptador con el sitio de unión molecular y los ácidos nucleicos de captura, que son lo suficientemente complementarios al sitio de unión molecular. Por lo tanto, a través de la "hibridación selectiva" se puede detectar la presencia del polinucleótido diana en una muestra impura que contiene un conjunto de muchos ácidos nucleicos. Un ejemplo de un sistema de aislamiento de hibridación selectiva nucleótido-nucleótido comprende un sistema con varios nucleótidos de captura, que son secuencias complementarias a los elementos de identificación de unión molecular, y opcionalmente se inmovilizan en un soporte sólido. En otras realizaciones, los polinucleótidos de captura podrían ser complementarios a las secuencias diana o un código de barras o etiqueta única contenida dentro del adaptador. Los polinucleótidos de captura se pueden inmovilizar en diversos soportes sólidos, como dentro de un pozo de una placa, esferas monodispersadas, microensayos o cualquier otra superficie de soporte adecuada conocida en la técnica. Los polinucleótidos adaptadores complementarios hibridados unidos en el soporte sólido se pueden aislar lavando los ácidos nucleicos no deseados no vinculantes, dejando atrás los polinucleótidos diana deseados. Si las moléculas adaptadoras complementarias se fijan a esferas paramagnéticas o tecnología de perlas similar para el aislamiento, las esferas se pueden mezclar en un tubo junto con el polinucleótido diana que contiene los adaptadores. Cuando las secuencias adaptadoras se han hibridado con las secuencias complementarias fijadas a las esferas, las moléculas indeseables se pueden lavar mientras las esferas se mantienen en el tubo con un imán o un agente similar. Las moléculas diana deseadas pueden liberarse posteriormente aumentando la temperatura, cambiando el pH o utilizando cualquier otro método de elución adecuado conocido en la técnica.

4. Códigos de barras

[0142] Un código de barras es una secuencia de ácido nucleico conocida que permite identificar alguna característica de un ácido nucleico con el que está asociado el código de barras. En algunas realizaciones, la característica del ácido nucleico a identificar es la muestra o fuente de la que se deriva el ácido nucleico. La secuencia de código de barras generalmente incluye ciertas características que hacen que la secuencia sea útil en las reacciones de secuenciación. Por ejemplo, las secuencias de códigos de barras están diseñadas para tener regiones de homopolímero mínimas o nulas, p. ej., 2 o más de la misma base en una fila, como AA o CCC, dentro de la secuencia de códigos de barras. En algunas realizaciones, las secuencias de códigos de barras también están diseñadas de modo que estén al menos a una distancia de edición del orden de adición de bases cuando se realiza una secuencia de base por base, asegurando que la primera y la última base no coincidan con las bases esperadas de la secuencia.

[0143] En algunas realizaciones, las secuencias de códigos de barras están diseñadas de tal manera que cada secuencia se correlaciona con un ácido nucleico diana particular, permitiendo que las lecturas de secuencia corta se correlacionen de nuevo con el ácido nucleico diana del que provienen. Los métodos para diseñar conjuntos de secuencias de códigos de barras se muestran, p. ej., en la patente de EE.UU. N° 6.235.475. En algunas realizaciones, las secuencias de códigos de barras varían de aproximadamente 5 nucleótidos a aproximadamente 15 nucleótidos. En una realización particular, las secuencias de código de barras varían de aproximadamente 4 nucleótidos a aproximadamente 7 nucleótidos. Dado que las secuencias de códigos de barras se secuencian junto con el fragmento de ácido nucleico en escalera, en las realizaciones que usan secuencias más largas, la longitud del código de barras es de una longitud mínima para permitir la lectura más larga del fragmento de ácido nucleico unido al código de barras. En algunas realizaciones, las secuencias de códigos de barras están separadas del fragmento de la molécula de ácido nucleico por al menos una base, p. ej., para minimizar las combinaciones homopoliméricas.

[0144] En algunas realizaciones, las longitudes y secuencias de secuencias de códigos de barras están diseñadas para lograr un nivel deseado de precisión para determinar la identidad del ácido nucleico. Por ejemplo, en algunas realizaciones, las secuencias de códigos de barras están diseñadas de tal manera que después de un número tolerable de mutaciones puntuales, la identidad del ácido nucleico asociado todavía se puede deducir con la precisión deseada. En algunas realizaciones, una transposasa de Tn-5 (disponible comercialmente en Epicenter Biotechnologies; Madison, Wisconsin) corta un ácido nucleico en fragmentos e inserta pedazos cortos de ADN en los cortes. Los fragmentos cortos de ADN se utilizan para incorporar las secuencias de códigos de barras.

[0145] La unión de adaptadores que comprenden códigos de barras a plantillas de ácido nucleico se muestra en la solicitud de patente de EE.UU. Pub. N° 2008/0081330 y en solicitud de patente internacional N° PCT/US09/64001. Los métodos para diseñar conjuntos de secuencias de códigos de barras y otros métodos para unir adaptadores (p. ej., que comprenden secuencias de códigos de barras) se muestran en la patente de EE.UU. N°s 6,138,077; 6,352,828; 5,636,400; 6,172,214; 6,235,475; 7,393,665; 7,544,473; 5,846,719; 5,695,934; 5,604,097; 6,150,516; RE39,793; 7,537,897; 6,172,218; y 5,863,722. En ciertas realizaciones, se une un único código de barras a cada fragmento. En otras realizaciones, se une una pluralidad de códigos de barras, p. ej., dos códigos de barras, a cada fragmento.

5. Muestras

[0146] En algunas realizaciones, las moléculas de plantilla de ácido nucleico (p. ej., ADN o ARN) se aíslan de una muestra biológica que contiene una variedad de otros componentes, tales como proteínas, lípidos y ácidos nucleicos no plantilla. Las moléculas de plantilla de ácido nucleico pueden obtenerse de cualquier material (p. ej., material celular (vivo o muerto), material extracelular, material viral, muestras ambientales (p. ej., muestras metagenómicas), material sintético (p. ej., amplicones como los proporcionados por PCR u otras tecnologías de amplificación)), obtenido de un

animal, planta, bacteria, arqueo, hongo o cualquier otro organismo. Las muestras biológicas para uso en la presente invención incluyen partículas virales o preparaciones de las mismas. Las moléculas de plantilla de ácido nucleico se pueden obtener directamente de un organismo o de una muestra biológica obtenida de un organismo, p. ej., de sangre, orina, líquido cefalorraquídeo, líquido seminal, saliva, esputo, heces, cabello, sudor, lágrimas, piel y tejido. Las muestras ejemplares incluyen, pero no se limitan a, sangre entera, líquido linfático, suero, plasma, células bucales, sudor, lágrimas, saliva, esputo, cabello, piel, biopsia, líquido cefalorraquídeo (LCR), líquido amniótico, líquido seminal, excreciones vaginales, líquido seroso, líquido sinovial, líquido pericárdico, líquido peritoneal, líquido pleural, transudados, exudados, líquido quístico, bilis, orina, fluidos gástricos, fluidos intestinales, muestras fecales y hisopos, aspirados (p. ej., médula ósea, aguja fina, etc.), lavados (p. ej., oral, nasofaríngeo, bronquial, bronquialveolar, óptico, rectal, intestinal, vaginal, epidérmico, etc.) y/u otras muestras.

[0147] Cualquier muestra de tejido o fluido corporal puede usarse como fuente de ácido nucleico para usar en la tecnología, incluyendo muestras forenses, muestras archivadas, muestras preservadas y/o muestras almacenadas durante largos períodos de tiempo, p. ej., congelados en fresco, y muestras fijadas en metanol/ácido acético o incluidas en parafina fijadas en formalina (FFPE). Las moléculas de plantilla de ácido nucleico también se pueden aislar de células cultivadas, como un cultivo celular primario o una línea celular. Las células o tejidos de los que se obtienen los ácidos nucleicos de plantilla pueden infectarse con un virus u otro patógeno intracelular. Una muestra también puede ser ARN total extraído de una muestra biológica, una biblioteca de ADNc, ADN viral o genómico. Una muestra también puede ser ADN aislado de un origen no celular, p. ej., ADN amplificado/aislado que se ha almacenado en un congelador.

[0148] Moléculas de plantilla de ácido nucleico se pueden obtener, p. ej., por extracción de una muestra biológica, p. ej., por una variedad de técnicas tales como las descritas por Maniatis, et al. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY (véase, p. ej., págs. 280-281).

[0149] En algunas realizaciones, la selección del tamaño de los ácidos nucleicos se realiza para eliminar fragmentos muy cortos o fragmentos muy largos. Los métodos adecuados para seleccionar un tamaño son conocidos en la técnica. En diversas realizaciones, el tamaño está limitado a 0,5, 1, 2, 3, 4, 5, 7, 10, 12, 15, 20, 25, 30, 50, 100 kb o más.

[0150] En diversas realizaciones, se amplifica un ácido nucleico. Se puede usar cualquier método de amplificación conocido en la técnica. Los ejemplos de técnicas de amplificación que se pueden utilizar incluyen, entre otros, PCR, PCR cuantitativa, PCR fluorescente cuantitativa (QF-PCR), PCR fluorescente multiplex (MF-PCR), PCR en tiempo real (RT-PCR), célula única PCR, PCR de polimorfismo de longitud de fragmento de restricción (PCR-RFLP), PCR de arranque en caliente, PCR anidada, PCR de colonia in situ, amplificación de círculo rodante in situ (RCA), PCR de puente, PCR de picotiter y PCR de emulsión. Otros métodos de amplificación adecuados incluyen la reacción en cadena de la ligasa (LCR), amplificación de la transcripción, replicación de secuencia autosostenida, amplificación selectiva de secuencias de polinucleótidos diana, reacción en cadena de la polimerasa cebada de secuencia consenso (CP-PCR), reacción en cadena de la polimerasa cebada arbitrariamente (AP-PCR), PCR cebada con oligonucleótidos degenerados (DOP-PCR) y amplificación de secuencia basada en ácido nucleico (NABSA). Otros métodos de amplificación que se pueden usar en el presente documento incluyen los descritos en la patente de EE.UU. N^{os} 5,242,794; 5,494,810; 4,988,617; y 6,582,938.

[0151] En algunas realizaciones, la reparación final se lleva a cabo para generar extremos de ácido nucleico fosforilados en extremo 5' con extremos romos usando kits comerciales, tales como los disponibles en Epicenter Biotechnologies (Madison, Wisconsin).

6. Secuenciación de ácido nucleico

[0152] En algunas realizaciones de la tecnología, se generan datos de secuencia de ácido nucleico. Diversas realizaciones de plataformas de secuenciación de ácido nucleico (p. ej., un secuenciador de ácido nucleico) incluyen componentes como se describe a continuación. Según diversas realizaciones, un instrumento de secuenciación incluye una unidad de administración y control de fluidos, una unidad de procesamiento de muestras, una unidad de detección de señal y una unidad de adquisición, análisis y control de datos. Diversas realizaciones del instrumento proporcionan una secuencia automatizada que se usa para recopilar información de secuencia de una pluralidad de secuencias en paralelo y/o sustancialmente simultáneamente.

[0153] En algunas realizaciones, la unidad de administración y control de fluidos incluye un sistema de administración de reactivos. El sistema de suministro de reactivo incluye un depósito de reactivo para el almacenamiento de diversos reactivos. Los reactivos pueden incluir cebadores basados en ARN, cebadores de ADN directo/inverso, mezclas de nucleótidos (p. ej., composiciones que comprenden análogos de nucleótidos como se proporciona aquí) para secuenciación por síntesis, tampones, reactivos de lavado, reactivos de bloqueo, reactivos de separación y similares. Además, el sistema de suministro de reactivo puede incluir un sistema de pipeteo o un sistema de flujo continuo que conecta la unidad de procesamiento de muestras con el depósito de reactivo.

[0154] En algunas realizaciones, la unidad de procesamiento de muestra incluye una cámara de muestra, tal como una célula de flujo, un sustrato, un microconjunto, una bandeja de múltiples pocillos o similares. La unidad de

procesamiento de muestras puede incluir múltiples carriles, múltiples canales, múltiples pozos u otros medios para procesar múltiples conjuntos de muestras de manera sustancialmente simultánea. Además, la unidad de procesamiento de muestras puede incluir múltiples cámaras de muestras para permitir el procesamiento de múltiples ejecuciones simultáneamente. En realizaciones particulares, el sistema puede realizar la detección de señal en una cámara de muestra mientras procesa de manera sustancial y simultánea otra cámara de muestra. Además, la unidad de procesamiento de muestras puede incluir un sistema de automatización para mover o manipular la cámara de muestras. En algunas realizaciones, la unidad de detección de señal puede incluir una imagen o sensor de detección. Por ejemplo, el sensor de imágenes o detección (p. ej., un detector de fluorescencia o un detector eléctrico) puede incluir un CCD, un CMOS, un sensor de iones, como una capa sensible a iones que recubre un CMOS, un detector de corriente o similar. La unidad de detección de señal puede incluir un sistema de excitación para hacer que una sonda, como un tinte fluorescente, emita una señal. El sistema de detección puede incluir una fuente de iluminación, como una lámpara de arco, un láser, un diodo emisor de luz (LED) o similar. En realizaciones particulares, la unidad de detección de señal incluye óptica para la transmisión de luz desde una fuente de iluminación a la muestra o desde la muestra al sensor de imagen o detección. Alternativamente, la unidad de detección de señal puede no incluir una fuente de iluminación, como p. ej., cuando se produce una señal espontáneamente como resultado de una reacción de secuenciación. Por ejemplo, se puede producir una señal por la interacción de un resto liberado, como un ion liberado que interactúa con una capa sensible a iones, o un pirofosfato que reacciona con una enzima u otro catalizador para producir una señal quimioluminiscente. En otro ejemplo, los cambios en una corriente eléctrica, voltaje o resistencia se detectan sin la necesidad de una fuente de iluminación.

[0155] En algunas realizaciones, una unidad de análisis y control de adquisición de datos monitorea varios parámetros del sistema. Los parámetros del sistema pueden incluir la temperatura de varias partes del instrumento, como la unidad de procesamiento de muestras o depósitos de reactivos, volúmenes de diversos reactivos, el estado de varios subcomponentes del sistema, como un manipulador, un motor paso a paso, una bomba o similares, o cualquier combinación de los mismos.

[0156] Un experto en la materia apreciará que se utilizan diversas realizaciones de los instrumentos y sistemas para practicar métodos de secuenciación tales como secuenciación por síntesis, métodos de molécula única y otras técnicas de secuenciación. La secuenciación por síntesis puede incluir la incorporación de nucleótidos marcados con colorante, terminación de cadena, secuenciación de iones/protones, secuenciación de pirofosfato o similares. Las técnicas de molécula única pueden incluir secuenciación escalonada, donde las reacciones de secuenciación se pausan para determinar la identidad del nucleótido incorporado.

[0157] En algunas realizaciones, el instrumento de secuenciación determina la secuencia de un ácido nucleico, tal como un polinucleótido o un oligonucleótido. El ácido nucleico puede incluir ADN o ARN, y puede ser monocatenario, como ADNss y ARN, o bicatenario, como ADNds o un par de ARN/ADNc. En algunas realizaciones, el ácido nucleico puede incluir o derivarse de una biblioteca de fragmentos, una biblioteca de pares de parejas, un fragmento de CHIP o similares. En realizaciones particulares, el instrumento de secuenciación puede obtener la información de secuencia de una sola molécula de ácido nucleico o de un grupo de moléculas de ácido nucleico sustancialmente idénticas.

[0158] En algunas realizaciones, el instrumento de secuenciación puede generar datos de lectura de secuenciación de ácido nucleico en una variedad de diferentes tipos/formatos de archivos de datos de salida, que incluyen, pero no se limitan a: *.txt, *.fasta, *.csfasta, * seq.txt, * qseq.txt, *.fastq, *.sff, * prb.txt, *.sms, * srs y/o *.qv.

7. Tecnologías de secuenciación de próxima generación

[0159] Las tecnologías de secuenciación particulares contempladas por la tecnología son los métodos de secuenciación de próxima generación (NGS) que comparten la característica común de estrategias de alto rendimiento masivamente paralelas, con el objetivo de reducir los costos en comparación con los más antiguos. métodos de secuenciación (véase, p. ej., Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296). Los métodos NGS se pueden dividir en términos generales en aquellos que generalmente usan amplificación de plantilla y aquellos que no. Los métodos que requieren amplificación incluyen pirosecuenciación comercializada por Roche como las plataformas tecnológicas 454 (p. ej., GS 20 y GS FLX), la plataforma Solexa comercializada por Illumina y la plataforma de detección y ligadura de oligonucleótidos compatibles (SOLiD) comercializada por Applied Biosystems. Los enfoques de no amplificación, también conocidos como secuenciación de una sola molécula, se ejemplifican con la plataforma HeliScope comercializada por Helicos BioSciences, y las plataformas emergentes comercializadas por VisiGen, Oxford Nanopore Technologies Ltd., Life Technologies/Ion Torrent y Pacific Biosciences, respectivamente.

[0160] En pirosecuenciación (Voelkerding y col., *Clinical Chem.*, 55: 641-658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287-296; Patente de Estados Unidos N° 6,210,891; Patente de EE.UU. 6,258,568), la biblioteca de fragmentos NGS se amplifica clonalmente in situ mediante la captura de moléculas de plantilla individuales con perlas que llevan oligonucleótidos complementarios a los adaptadores. Cada cuenta que lleva un tipo de plantilla única se divide en compartimentos en una microvesícula de agua en aceite, y la plantilla se amplifica clonalmente usando una técnica denominada PCR en emulsión. La emulsión se rompe después de la amplificación y las perlas se depositan en los pocillos individuales de una placa de picotitro que funciona como una célula de flujo durante las reacciones de

secuenciación. La introducción ordenada e iterativa de cada uno de los cuatro reactivos dNTP se produce en la célula de flujo en presencia de enzimas de secuenciación y reportero luminiscente como la luciferasa. En el caso de que se agregue un dNTP apropiado al extremo 3' de la secuencia, la producción resultante de ATP provoca una explosión de luminiscencia dentro del pozo, que se graba con una cámara CCD. Es posible lograr longitudes de lectura mayores o iguales a 400 bases, y se pueden lograr 10^6 lecturas de secuencia, lo que resulta en hasta 500 millones de pares de bases (Mb) de secuencia.

[0161] En la plataforma Solexa/Illumina (Voelkerding y col., *Clinical Chem.*, 55: 641-658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287-296; Patente de los Estados Unidos Núm. 6,833,246; Patente de Estados Unidos N° 7,115,400; Patente de Estados Unidos N° 6,969,488), los datos de secuenciación se producen en forma de lecturas de menor longitud. En este método, los fragmentos de la biblioteca de fragmentos NGS se capturan en la superficie de una célula de flujo que está tachonada con anclajes de oligonucleótidos. El anclaje se usa como cebador de PCR, pero debido a la longitud de la plantilla y su proximidad a otros oligonucleótidos de anclaje cercanos, la extensión por PCR da como resultado el "arqueamiento" de la molécula para hibridarse con un oligonucleótido de anclaje adyacente para formar una estructura de puente en la superficie de la célula de flujo. Estos bucles de ADN están desnaturalizados y cortados. Los filamentos hacia adelante se secuencian luego con terminadores de tinte reversibles. La secuencia de nucleótidos incorporados se determina mediante la detección de fluorescencia posterior a la incorporación, eliminando cada flúor y bloque antes del próximo ciclo de adición de dNTP. La longitud de lectura de secuencia varía de 36 nucleótidos a más de 100 nucleótidos, con una producción total que supera los mil millones de pares de nucleótidos por ciclo analítico.

[0162] La secuenciación de moléculas de ácido nucleico utilizando la tecnología de sólido (Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean y otros, *Nature Rev. Microbiol.*, 7: 287-296; Patente de Estados Unidos N° 5,912,148; la Patente de Estados Unidos N° 6,130,073) también implica la amplificación clonal de la biblioteca de fragmentos de NGS por PCR en emulsión. Después de esto, las perlas que llevan la plantilla se inmovilizan en una superficie derivatizada de una célula de flujo de vidrio, y se recoce un cebador complementario al oligonucleótido adaptador. Sin embargo, en lugar de utilizar este cebador para la extensión 3', se usa para proporcionar un grupo fosfato 5' para la ligadura a sondas de interrogación que contienen dos bases específicas de la sonda seguidas de 6 bases degeneradas y una de las cuatro etiquetas fluorescentes. En el sistema SOLiD, las sondas de interrogación tienen 16 combinaciones posibles de las dos bases en el extremo 3' de cada sonda, y uno de los cuatro fluoruros en el extremo 5'. El color flúor, y por lo tanto la identidad de cada sonda, corresponde a los esquemas de codificación de espacio de color especificados. Múltiples rondas (generalmente 7) de recocido de sonda, ligadura y detección de flúor son seguidas por desnaturalización, y luego una segunda ronda de secuenciación usando un cebador que está compensado por una base con respecto al cebador inicial. De esta manera, la secuencia de la plantilla se puede reconstruir computacionalmente, y las bases de la plantilla se interrogan dos veces, lo que resulta en una mayor precisión. La longitud de lectura de secuencia promedia 35 nucleótidos, y la producción total excede los 4 mil millones de bases por secuenciación.

[0163] En ciertas realizaciones, se emplea HeliScope de Helicos BioSciences (Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; Patente de Estados Unidos N° 7,169,560; Patente de Estados Unidos N° 7,282,337; Patente de Estados Unidos N° 7,482,120; Patente de Estados Unidos N° 7,501,245; Patente de Estados Unidos N° 6,818,395; Patente de Estados Unidos N° 6,911,345; Patente de Estados Unidos N° 7,501,245). La secuenciación se logra mediante la adición de polimerasa y la adición en serie de reactivos dNTP marcados con fluorescencia. Los eventos de incorporación dan como resultado una señal de fluor correspondiente al dNTP, y la señal es capturada por una cámara CCD antes de cada ronda de adición de dNTP. La longitud de lectura de secuencia varía de 25 a 50 nucleótidos, con una producción total superior a mil millones de pares de nucleótidos por análisis.

[0164] En algunas realizaciones, se usa la secuenciación 454 de Roche (Margulies et al. (2005) *Nature* 437: 376-380). La secuenciación 454 implica dos pasos. En el primer paso, el ADN se corta en fragmentos de aproximadamente 300-800 pares de bases y los fragmentos tienen extremos romos. Los adaptadores de oligonucleótidos se unen luego a los extremos de los fragmentos. Los adaptadores sirven como cebadores para la amplificación y secuenciación de los fragmentos. Los fragmentos se pueden unir a perlas de captura de ADN, p. ej., perlas recubiertas con estreptavidina usando, p. ej., un adaptador que contiene una etiqueta de 5'-biotina. Los fragmentos unidos a las perlas se amplifican por PCR dentro de las gotas de una emulsión de aceite y agua. El resultado son múltiples copias de fragmentos de ADN amplificados clonalmente en cada cuenta. En el segundo paso, las cuentas se capturan en pozos (tamaño picolitro). La pirosecuenciación se realiza en cada fragmento de ADN en paralelo. La adición de uno o más nucleótidos genera una señal de luz que es grabada por una cámara CCD en un instrumento de secuenciación. La intensidad de la señal es proporcional al número de nucleótidos incorporados. La pirosecuenciación hace uso de pirofosfato (PPi) que se libera tras la adición de nucleótidos. PPi se convierte en ATP por ATP sulfurilasa en presencia de adenosina 5' fosfosulfato. Luciferase usa ATP para convertir la luciferina en oxiluciferina, y esta reacción genera luz que se detecta y analiza.

[0165] La tecnología de torrente de iones es un método de secuenciación de ADN basado en la detección de iones de hidrógeno que se liberan durante la polimerización de ADN (véase, p. ej., *Science* 327 (5970): 1190 (2010); Sol. de Pat. de EE.UU. Pub. N°s 20090026082, 20090127589, 20100301398, 20100197507, 20100188073 y 20100137143).

Un micropocillo contiene un fragmento de la biblioteca de fragmentos NGS para ser secuenciado. Debajo de la capa de micropocillos hay un sensor de iones ISFET hipersensible. Todas las capas están contenidas dentro de un chip semiconductor CMOS, similar al utilizado en la industria electrónica. Cuando se incorpora un dNTP en la cadena complementaria en crecimiento, se libera un ion de hidrógeno, que activa un sensor de iones hipersensibles. Si las repeticiones de homopolímero están presentes en la secuencia de plantilla, se incorporarán múltiples moléculas de dNTP en un solo ciclo. Esto conduce a un número correspondiente de hidrógenos liberados y una señal electrónica proporcionalmente más alta. Esta tecnología difiere de otras tecnologías de secuenciación en que no se utilizan nucleótidos u ópticos modificados. La precisión por base del secuenciador de torrente de iones es ~ 99,6% para 50 lecturas base, con ~ 100 Mb generados por ejecución. La longitud de lectura es de 100 pares de bases. La precisión para las repeticiones de homopolímero de 5 repeticiones de longitud es ~ 98%. Los beneficios de la secuenciación de semiconductores iónicos son la velocidad de secuenciación rápida y los bajos costos iniciales y operativos. Sin embargo, el costo de adquisición de un secuenciador mediado por pH es de aproximadamente 50.000 \$, excluyendo el equipo de preparación de muestras y un servidor para el análisis de datos.

[0166] Stratos Genomics, Inc. desarrolló otro enfoque de secuenciación de ácido nucleico ejemplar que se puede adaptar para usar con la presente invención e implica el uso de Xpandomeros. Este proceso de secuenciación típicamente incluye proporcionar una cadena hija producida por una síntesis dirigida por plantilla. La cadena hija generalmente incluye una pluralidad de subunidades acopladas en una secuencia correspondiente a una secuencia de nucleótidos contigua de todo o una porción de un ácido nucleico diana en donde las subunidades individuales comprenden una correa, al menos una sonda o residuo de nucleobase, y al menos un enlace selectivamente escindible. El enlace o enlaces selectivamente escindibles se cortan para producir un Xpandomero de una longitud mayor que la pluralidad de las subunidades de la cadena hija. El Xpandomero típicamente incluye las ataduras y los elementos informadores para analizar la información genética en una secuencia correspondiente a la secuencia contigua de nucleótidos de todo o una porción del ácido nucleico diana. Luego se detectan elementos informadores del Xpandomero. Detalles adicionales relacionados con los enfoques basados en Xpandomeros se describen, p. ej., en la patente de EE.UU. Pub. 20090035777, titulada "SECUENCIACIÓN DE ÁCIDO NUCLEICO DE ALTO PASO POR EXPANSIÓN", presentada el 19 de junio de 2008.

[0167] Otros métodos de secuenciación de moléculas individuales incluyen secuenciación en tiempo real mediante síntesis utilizando una plataforma VisiGen (Voelkerding et al., *Clinical Chem.*, 55: 641-58, 2009; Patente de Estados Unidos N° 7,329,492; Solicitud de Patente de Estados Unidos N° de serie 11/671956; Solicitud de Patente de Estados Unidos N° de serie 11/781166) en donde fragmentos de la biblioteca de fragmentos NGS se inmovilizan, ceban, luego se someten a extensión de cadena usando una polimerasa modificada fluorescentemente y moléculas aceptoras fluorescentes, dando como resultado una transferencia de energía de resonancia de fluorescencia detectable (FRET) tras la adición de nucleótidos.

[0168] Otro sistema de secuenciación de moléculas individuales en tiempo real desarrollado por Pacific Biosciences (Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; Patente de Estados Unidos N° 7,170,050; Patente de Estados Unidos N° 7,302,146; Patente de Estados Unidos N° 7,313,308; patente de EE.UU. N° 7,476,503) utiliza pocillos de reacción de 50-100 nm de diámetro y que abarca un volumen de reacción de aproximadamente 20 zeptolitros (10^{-21} l). Las reacciones de secuenciación se realizan usando una plantilla inmovilizada, polimerasa de ADN phi29 modificada y altas concentraciones locales de dNTP marcados con fluorescencia. Las altas concentraciones locales y las condiciones de reacción continua permiten capturar los eventos de incorporación en tiempo real mediante la detección de la señal de flúor utilizando excitación láser, una guía de ondas ópticas y una cámara CCD.

[0169] En ciertas realizaciones, se emplean los métodos de secuenciación de ADN en tiempo real de una sola molécula (SMRT) usando guías de onda de modo cero (ZMW) desarrolladas por Pacific Biosciences, o métodos similares. Con esta tecnología, la secuenciación de ADN se realiza en chips SMRT, cada uno con miles de guías de onda de modo cero (ZMW). Un ZMW es un orificio, de decenas de nanómetros de diámetro, fabricado en una película metálica de 100 nm depositada sobre un sustrato de dióxido de silicio. Cada ZMW se convierte en una cámara de visualización nanofotónica proporcionando un volumen de detección de sólo 20 zeptolitros (10^{-21} l). En este volumen, la actividad de una sola molécula se puede detectar en un fondo de miles de nucleótidos marcados. El ZMW proporciona una ventana para observar la polimerasa de ADN mientras realiza la secuenciación por síntesis. Dentro de cada cámara, una sola molécula de polimerasa de ADN está unida a la superficie inferior de manera que permanezca permanentemente dentro del volumen de detección. Los nucleótidos fosforilados, cada tipo marcado con un fluoróforo de diferente color, luego se introducen en la solución de reacción a altas concentraciones que promueven la velocidad, precisión y procesabilidad de las enzimas. Debido al pequeño tamaño de la ZMW, incluso a estas altas concentraciones biológicamente relevantes, el volumen de detección está ocupado por nucleótidos solo una pequeña fracción del tiempo. Además, las visitas al volumen de detección son rápidas, duran solo unos pocos microsegundos, debido a la distancia muy pequeña que tiene la difusión para transportar los nucleótidos. El resultado es un fondo muy bajo.

[0170] En algunas realizaciones, se usa la secuenciación de nanoporos (Soni GV y Meller A. (2007) *Clin Chem* 53: 1996-2001). Un nanoporo es un pequeño agujero, del orden de 1 nanómetro de diámetro. La inmersión de un nanoporo en un fluido conductor y la aplicación de un potencial a través de él da como resultado una ligera corriente eléctrica

debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño del nanoporo. Cuando una molécula de ADN pasa a través de un nanoporo, cada nucleótido en la molécula de ADN obstruye el nanoporo en un grado diferente. Por lo tanto, el cambio en la corriente que pasa a través del nanoporo a medida que la molécula de ADN pasa a través del nanoporo representa una lectura de la secuencia de ADN.

[0171] En algunas realizaciones, una técnica de secuenciación utiliza una matriz de transistor de efecto de campo sensible a químicos (chemFET) para secuenciar ADN (p. ej., como se describe en la Publicación de Solicitud de Patente de Estados Unidos N° 20090026082). En un ejemplo de la técnica, las moléculas de ADN se colocan en cámaras de reacción, y las moléculas de plantilla se hibridan con un cebador de secuenciación unido a una polimerasa. La incorporación de uno o más trifosfatos en una nueva cadena de ácido nucleico en el extremo 3' del cebador de secuenciación puede detectarse mediante un cambio en la corriente por un chemFET. Una matriz puede tener múltiples sensores chemFET. En otro ejemplo, los ácidos nucleicos individuales pueden unirse a las perlas, y los ácidos nucleicos pueden amplificarse en la perla, y las perlas individuales pueden transferirse a cámaras de reacción individuales en una matriz de chemFET, con cada cámara con un sensor de chemFET, y los ácidos nucleicos pueden ser secuenciados.

[0172] En algunas realizaciones, la técnica de secuenciación utiliza un microscopio electrónico (Moudrianakis EN y Beer M. Proc Natl Acad Sci EE.UU. 1965 marzo; 53: 564-71). En un ejemplo de la técnica, las moléculas de ADN individuales se marcan con etiquetas metálicas que se distinguen con un microscopio electrónico. Estas moléculas luego se estiran sobre una superficie plana y se toman imágenes usando un microscopio electrónico para medir secuencias.

[0173] En algunas realizaciones, "secuenciación de cuatro colores por síntesis usando terminadores reversibles de nucleótidos fluorescentes escindibles" como se describe en Turro, et al. PNAS 103: 19635-40 (2006) se utiliza, p. ej., como comercializado por Intelligent Bio-Systems. La tecnología descrita en la publicación de solicitud de patente de EE.UU. N°s 2010/0323350, 2010/0063743, 2010/0159531, 20100035253, 20100152050.

[0174] Los procesos y sistemas para tal secuenciación en tiempo real que pueden adaptarse para su uso con la invención se describen, p. ej., en la patente de EE.UU. N° 7,405,281, titulada "Análogos de nucleótidos fluorescentes y usos para los mismos", publicada el 29 de julio de 2008 a Xu et al.; 7,315,019, titulada "Matrices de confinamientos ópticos y usos de las mismas", publicada de enero de 2008 a Turner et al.; 7,313,308, titulada "Análisis óptico de moléculas", publicada el 25 de diciembre de 2007 a Turner et al.; 7,302,146, titulada "Aparato y método para el análisis de moléculas", publicada el 27 de noviembre de 2007 a Turner et al.; y 7,170,050, titulada "Aparatos y métodos para el análisis óptico de moléculas", publicada el 30 de enero de 2007 a Turner et al.; y la patente de EE.UU. Pub. N°s 20080212960, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 26 de octubre de 2007 por Lundquist et al.; 20080206764, titulada "Sistema Flowcell para la detección de moléculas individuales", presentada el 26 de octubre de 2007 por Williams et al.; 20080199932, titulada "Polimerasas de superficie activa acopladas", presentada el 26 de octubre de 2007 por Hanzel et al.; 20080199874, titulada "ESCISIÓN DE HILO CONTROLABLE DE ADN DE MINI CÍRCULO", presentada el 11 de febrero de 2008 por Otto et al.; 20080176769, titulada "Artículos que tienen moléculas localizadas dispuestas sobre ellos y métodos para producirlos", presentada el 26 de octubre de 2007 por Rank et al.; 20080176316, titulada "Mitigación del fotodaño en reacciones analíticas", presentada el 31 de octubre de 2007 por Eid et al.; 20080176241, titulada "Mitigación del fotodaño en reacciones analíticas", presentada el 31 de octubre de 2007 por Eid et al.; 20080165346, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 26 de octubre de 2007 por Lundquist et al.; 20080160531, titulada "Superficies uniformes para sustratos de materiales híbridos y métodos para hacer y usar los mismos", presentada el 31 de octubre de 2007 por Korch; 20080157005, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 26 de octubre de 2007 por Lundquist et al.; 20080153100, titulada "Artículos que tienen moléculas localizadas dispuestas sobre ellos y métodos para producirlos", presentada el 31 de octubre de 2007 por Rank et al.; 20080153095, titulada "NUCLEOTIDOS DE INTERRUPTOR DE CARGA", presentada el 26 de octubre de 2007 por Williams et al.; 20080152281, titulada "Sustratos, sistemas y métodos para analizar materiales", presentada el 31 de octubre de 2007 por Lundquist et al.; 20080152280, titulada "Sustratos, sistemas y métodos para analizar materiales", presentada el 31 de octubre de 2007 por Lundquist et al.; 20080145278, titulada "Superficies uniformes para sustratos de materiales híbridos y métodos para hacer y usar los mismos", presentada el 31 de octubre de 2007 por Korch; 20080128627, titulada "SUSTRATOS, SISTEMAS Y MÉTODOS PARA ANALIZAR MATERIALES", presentada el 31 de agosto de 2007 por Lundquist et al.; 20080108082, titulada "Enzimas y reactivos de polimerasa para la secuenciación mejorada de ácidos nucleicos", presentada el 22 de octubre de 2007 por Rank et al.; 20080095488, titulada "SUSTRATOS PARA REALIZAR REACCIONES ANALÍTICAS", presentada el 11 de junio de 2007 por Foquet et al.; 20080080059, titulado "COMPONENTES ÓPTICOS MODULARES Y SISTEMAS QUE INCORPORAN LOS MISMOS", presentada el 27 de septiembre de 2007 por Dixon et al.; 20080050747, titulada "Artículos que tienen moléculas localizadas dispuestas sobre ellos y métodos de producción y uso de los mismos", presentada el 14 de agosto de 2007 por Korch et al.; 20080032301, titulada "Artículos que tienen moléculas localizadas dispuestas sobre ellos y métodos para producirlos", presentada el 29 de marzo de 2007 por Rank et al.; 20080030628, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 9 de febrero de 2007 por Lundquist et al.; 20080009007, titulada "INICIACIÓN CONTROLADA DE LA EXTENSIÓN DEL CEBADOR", presentada el 15 de junio de 2007 por Lyle et al.; 20070238679,

titulada "Artículos que tienen moléculas localizadas dispuestas sobre ellos y métodos para producirlos", presentada el 30 de marzo de 2006 por Rank et al.; 20070231804, titulada "Métodos, sistemas y composiciones para controlar la actividad enzimática y sus aplicaciones", presentada el 31 de marzo de 2006 por Korlach et al.; 20070206187, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 9 de febrero de 2007 por Lundquist et al.; 20070196846, titulada "Polimerasas para la incorporación de análogos de nucleótidos", presentada el 21 de diciembre de 2006 por Hanzel et al.; 20070188750, titulada "Métodos y sistemas para el monitoreo simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 7 de julio de 2006 por Lundquist et al.; 20070161017, titulada "MITIGACIÓN DE FOTODAÑO EN REACCIONES ANALÍTICAS", presentada el 1 de diciembre de 2006 por Eid et al.; 20070141598, titulada "Composiciones de nucleótidos y usos de las mismas", presentada el 3 de noviembre de 2006 por Turner et al.; 20070134128, titulada "Superficies uniformes para sustrato de material híbrido y métodos para fabricar y usar el mismo", presentado el 27 de noviembre de 2006 por Korlach; 20070128133, titulada "Mitigación del fotodaño en reacciones analíticas", presentada el 2 de diciembre de 2005 por Eid et al.; 20070077564, titulada "Superficies reactivas, sustratos y métodos de producción de los mismos", presentada el 30 de septiembre de 2005 por Roitman et al.; 20070072196, titulada "Análogos de nucleótidos fluorescentes y usos para los mismos", presentada el 29 de septiembre de 2005 por Xu et al., y 20070036511, titulada "Métodos y sistemas para monitorear múltiples señales ópticas de una sola fuente", presentada el 11 de agosto de 2005 por Lundquist et al.; y Korlach et al. (2008) "Pasivación selectiva de aluminio para la inmovilización dirigida de moléculas de polimerasa de ADN individuales en nanoestructuras de guía de onda de modo cero" PNAS 105 (4): 1176-81.

8. Análisis de secuencia de ácido nucleico

[0175] En alguna forma de realización, un programa de análisis basado en computadora se usa para traducir los datos brutos generados por el ensayo de detección (p. ej., lecturas de secuenciación) en datos de valor predictivo para un usuario final (p. ej., personal médico). El usuario puede acceder a los datos predictivos utilizando cualquier medio adecuado. Por lo tanto, en algunas realizaciones preferidas, la presente tecnología proporciona el beneficio adicional de que el usuario, que probablemente no esté capacitado en genética o biología molecular, no necesita comprender los datos en bruto. Los datos se presentan directamente al usuario final en su forma más útil. El usuario puede utilizar inmediatamente la información para determinar información útil (p. ej., en diagnósticos médicos, investigaciones o exámenes de detección).

[0176] La presente tecnología también abarca cualquier método capaz de recibir, procesar y transmitir la información hacia y desde los laboratorios que realizan los ensayos, la información proporcionada, el personal médico y los sujetos.

9. Usos

[0177] La tecnología no se limita a usos particulares, sino que se utiliza en una amplia gama de aplicaciones de investigación (básicas y aplicadas), clínicas, médicas y otras aplicaciones biológicas, bioquímicas y biológicas moleculares. Algunos usos ejemplares de la tecnología incluyen genética, genómica y/o genotipado, p. ej., de plantas, animales y otros organismos, p. ej., para identificar haplotipos, fases y/o enlaces de mutaciones y/o alelos. Ejemplos ilustrativos particulares y no limitativos en el contexto médico humano incluyen pruebas de fibrosis quística y síndrome de X frágil.

[0178] Además, la tecnología encuentra uso en el campo de las enfermedades infecciosas, p. ej., en la identificación de agentes infecciosos tales como virus, bacterias, hongos, etc., y en la determinación de tipos, familias, especies y/o cuasiespecies virales, e identificar haplotipos, fases y/o enlaces de mutaciones y/o alelos. Un ejemplo ilustrativo particular y no limitante en el área de las enfermedades infecciosas es la caracterización de elementos genéticos del virus de inmunodeficiencia humana (VIH) y la identificación de haplotipos, fases y/o enlaces de mutaciones y/o alelos. Otros ejemplos ilustrativos particulares y no limitativos en el área de enfermedades infecciosas incluyen la caracterización de determinantes de resistencia a antibióticos; seguimiento de organismos infecciosos para epidemiología; monitorear la aparición y evolución de mecanismos de resistencia; identificación de especies, subespecies, cepas, elementos extracromosómicos, tipos, etc. asociados con la virulencia, monitorizando el progreso de los tratamientos, etc.

[0179] En algunas realizaciones, la tecnología se utiliza en la medicina de trasplantes, p. ej., para la tipificación del complejo principal de histocompatibilidad (MHC), la tipificación del antígeno leucocitario humano (HLA), y para identificar haplotipos, fases y/o enlaces de mutaciones y/o alelos asociados con la medicina de trasplante (p. ej., para identificar donantes compatibles para un huésped particular que necesita un trasplante, para predecir la posibilidad de rechazo, para monitorear el rechazo, para archivar material de trasplante, para bases de datos de informática médica, etc.).

[0180] En algunas realizaciones, la tecnología encuentra uso en oncología y campos relacionados con la oncología. Ejemplos ilustrativos particulares y no limitantes en el área de la oncología son la identificación de aberraciones genéticas y/o genómicas relacionadas con el cáncer, la predisposición al cáncer y/o el tratamiento del cáncer. Por ejemplo, en algunas realizaciones, la tecnología encuentra uso en la detección de la presencia de una translocación cromosómica asociada con el cáncer; y en algunas realizaciones, la tecnología encuentra uso en la identificación de

nuevos pares de fusión génica para proporcionar pruebas de diagnóstico de cáncer. En algunas realizaciones, la tecnología encuentra uso en la detección del cáncer, el diagnóstico del cáncer, el pronóstico del cáncer, la medición de la enfermedad residual mínima y la selección y/o monitoreo de un curso de tratamiento para un cáncer.

5 **[0181]** En algunas realizaciones, la tecnología encuentra uso en la caracterización de secuencias de nucleótidos. Por ejemplo, en algunas realizaciones, la tecnología encuentra uso en la detección de inserciones y/o deleciones ("indeles") en una secuencia de nucleótidos (p. ej., genoma, gen, etc.). Se contempla que la tecnología descrita en el presente documento proporciona una detección indel mejorada en relación con las tecnologías convencionales. Además, la tecnología encuentra uso en la detección de repeticiones en tándem cortas (STR), inversiones, inserciones grandes y en la secuenciación de regiones repetitivas (p. ej., altamente repetitivas) de una secuencia de nucleótidos (p. ej., de un genoma).

10 **[0182]** Aunque la descripción en el presente documento se refiere a ciertas realizaciones ilustradas, debe entenderse que estas realizaciones se presentan a modo de ejemplo y no a modo de limitación.

15 **Ejemplos**

Ejemplo 1 - Comparación con Illumina MiSeq

20 **[0183]** Durante el desarrollo de la tecnología proporcionada en este documento, se realizaron cálculos para comparar el rendimiento de la tecnología proporcionada en este documento (Tablas 1 y 2, "Biblioteca SOD") con la tecnología convencional proporcionada por Illumina en la plataforma MiSeq (Tablas 1 y 2, "Biblioteca Illumina Amplicon"). Los datos se recopilaron para dos escenarios que varían, p. ej., el número de muestras por ejecución, los criterios para medir el rendimiento, etc. (ver las Tablas 1 y 2).

25 **[0184]** Como se muestra en las Tablas 1 y 2, la tecnología descrita en este documento disminuye el tiempo de ejecución del instrumento, tiene un rendimiento más alto y produce un mayor porcentaje de lecturas con puntuaciones de calidad superiores a Q30 con respecto a la construcción de la biblioteca NGS utilizando la tecnología Illumina.

30 **Tabla 1 - Comparación con Illumina MiSeq (secuenciación dirigida: panel de amplicón)**

MiSeq (kit de reactivos de secuenciación v2)^a		
	Biblioteca SOD	Biblioteca Illumina Amplicon
Nº de muestras por ejecución	8	8
Nº de amplicones por muestra	50	50
Tamaño promedio de amplicones (pb)	400	400
Longitud requerida de lectura de SBS	1 x 50	2 x 250 ^b
Tiempo total de ejecución (horas) ^c	3	37
Media cobertura para cada amplicón por muestra ^d	5357	37500
Rendimiento (nº de muestras con cobertura 1000x/hora) ^e	14,3	8,1
Puntajes de calidad (porcentaje de lecturas con puntaje > Q30) ^f	> 90%	> 75%
a) Kit de reactivos MiSeq v2: escaneo de doble superficie, 12-15 millones de clústeres que pasan el filtro b) Para cubrir todo el amplicón de 400 pb, se implementa una estrategia de lectura de 2 x 250 pb en el extremo del par donde las lecturas se superponen por < 100 pb c) Solo la porción de secuencia real (no incluye el tiempo de generación del clúster) d) Para calcular la cobertura de la biblioteca SOD: $[(n^{\circ} \text{ total de lecturas}) / ((\text{tamaño de inserción} - \text{longitud de lectura SOD}) \times (n^{\circ} \text{ de muestras en una ejecución} \times n^{\circ} \text{ de amplicones por muestra}))] \times \text{SOD longitud de lectura}$: p. ej., $[(15 \times 10^6) / ((400 - 50) \times (8 \times 50))] \times 50$ e) Para calcular el rendimiento: $[(\text{cobertura media}) / 1000] / (\text{tiempo total de ejecución})$ f) Basado en la especificación de secuencia de MiSeq proporcionada por Illumina, p. ej., en sus materiales en línea.		

Tabla 2: Comparación con Illumina MiSeq (secuenciación de panel dirigida de inserto de 400 bp)

MiSeq (seq kit v2)			
	Secuenciación de panel dirigida (inserto de 400 pb)		
	Biblioteca SOD	Biblioteca Amplicon	Illumina
N.º de muestras por análisis	8	8	56
N.º de <u>amplicones</u> por muestra	50	50	50
Tamaño promedio de amplicones (bp)	400	400	400
Longitud requerida de lectura de SBS	1 x 50	2 x 250 ^b	2 x 250 ^b
Tiempo total de ejecución (horas)^c	4	38	38
Cobertura media para cada amplicón por muestra ^d	5357	37500	5357
Rendimiento (nº de muestras con 2000x cobertura/hora) ^e	7,1	4,1	4,1
Puntajes de calidad (porcentaje de lecturas con puntaje > Q30) ^f	>90%	>75%	>75%
a) Kit de reactivos MiSeq v2: escaneo de doble superficie, 15 millones de clústeres que pasan el filtro b) Para cubrir todo el amplicón de 200 o 400 pb, se implementa una estrategia de lectura de extremo de par en donde las lecturas se solapan por ~ 100 bp c) Parte de secuenciación real solamente (no incluye el tiempo de generación de clúster) d) Para la cobertura de cálculo para la biblioteca SOD: $[(n^{\circ} \text{ total de lecturas}) / ((\text{tamaño de inserto} - \text{longitud de lectura SOD}) \times (n^{\circ} \text{ de muestras en una serie} \times n^{\circ} \text{ de amplicones por muestra}))] \times \text{SOD}$ longitud de lectura: p. ej., $[(15 \times 10^6) / ((400 - 50) \times (8 \times 50))] \times 50$ e) Para calcular el rendimiento: $[(\text{cobertura media}) / 2000] / (\text{tiempo total de ejecución})$ f) Basado en la especificación de secuencia MiSeq proporcionada por Illumina, p. ej., en sus materiales en línea.			

Ejemplo 2 - Comparación con Torrente de Iones PGM (secuenciación dirigida: panel de amplicón)

[0185] Durante el desarrollo de la tecnología proporcionada en este documento, se realizaron cálculos para comparar el rendimiento de la tecnología proporcionada aquí (Tablas 3 y 4, "Biblioteca SOD") con tecnología convencional proporcionada por Torrente de Iones en la plataforma PGM (Tablas 3 y 4, "Biblioteca de Amplicon de Iones"). Los datos se recopilaron para dos escenarios que varían, p. ej., el número de muestras por ejecución, los criterios para medir el rendimiento, etc. (ver las Tablas 3 y 4).

[0186] Como se muestra en las Tablas 3 y 4, la tecnología descrita en el presente documento disminuye el tiempo de ejecución del instrumento y produce un mayor porcentaje de lecturas con puntuaciones de calidad superiores a Q20 con respecto a la construcción de la biblioteca NGS utilizando la tecnología de Torrente de Iones.

Tabla 3: Comparación con Ion Torrent PGM

Ion PGM (kit de reactivo de secuenciación de 400 pb v2) ^a		
	Biblioteca SOD	Biblioteca de Amplicon de Iones
Nº de muestras por ejecución	1	1
Nº de <u>amplicones</u> por muestra	50	50
Tamaño promedio de amplicones (bp)	400	400
Longitud requerida de lectura de SBS	1 x 50	1 x 400 (bidireccional) ^b
Tiempo total de ejecución (horas) ^c	0,5	4
Media cobertura para cada amplicón por muestra ^d	1143	8000
Rendimiento (nº de muestras con cobertura 1000x/hora) ^e	2,3	2,0
Puntajes de calidad (porcentaje de lecturas con puntaje >Q20) ^f	> 90%	> 50%
a) Kit de reactivo de secuenciación PGM de 400 pb v2 b) Para cubrir todo el amplicón de 400 pb, se realiza una secuenciación bidireccional de 1 X 400 pb c) Solo una porción de secuencia real (no incluye OneTouch2 y otro tiempo de proceso de pre-secuenciación) d) Para calcular la cobertura de la biblioteca SOD: $[(0,4 \times 10^6) / ((400-50) \times (8 \times 50))] \times 50$ e) Para calcular el rendimiento: $[(\text{cobertura promedio}) / 1000] / (\text{total tiempo de ejecución})$ f) Proyectado en base a: Loman N. et al. (2012) "Comparación de rendimiento de plataformas de secuenciación de alto rendimiento de sobremesa " Nature Biotechnology, vol. 30-5.		

Tabla 4 -Comparación con Torrente de Iones PGM

Ion PGM (kit 318 v2/seq) ^a			
Secuenciación de panel dirigida (inserto de 400 pb)			
	Biblioteca SOD	Biblioteca de Amplicon de Iones	
Nº de muestras por ejecución	4	4	28
Nº de amplicones por muestra	50	50	50
Tamaño promedio de amplicones (bp)	400	400	400
Longitud requerida de lectura de SBS	1 x 50	1 x 400 (bidireccional) ^b	
Tiempo total de ejecución (horas) ^c	0,5	7,25	7,25
Media cobertura para cada amplicón por muestra ^d	4286	30000	4286
Rendimiento (nº de muestras con cobertura 2000x/hora) ^e	17,1	8,3	8,3
Puntajes de calidad (porcentaje de lecturas con puntaje >Q20) ^f	> 90%	> 50%	> 50%
a) Ion PGM chip 318/v2: ~ 6 millones de pozos de carga que producen lecturas que pasan el filtro b) Para cubrir todo el amplicón de 200 bp o 400 bp, un 200 bp (bidireccional) o 400 bp (bidireccional) se implementa la estrategia, respectivamente c) Solo la porción de secuencia real (no incluye ePCR/enriquecimiento) d) Para calcular la cobertura de la biblioteca SOD:[(nº de lecturas totales)/((tamaño de inserción - longitud de lectura SOD) X (nº de muestras X nº de amplicón))] X SOD longitud de lectura, p. ej., [(15 X 10 ⁶)/((400 - 50) X (8 X 50))] X 50 e) Para calcular el rendimiento:[(cobertura media)/2000]/(tiempo de ejecución total) f) Según la especificación de secuenciación de Ion Torrent disponible en los materiales en línea de torrente de iones			

Ejemplo 3 - Tecnologías de comparación para lecturas largas

[0187] Las Tablas 5 y 6 comparan el rendimiento de la tecnología proporcionada aquí con tecnologías convencionales para secuenciar amplicones largos de aproximadamente 1000 bp (Tabla 5) y 2000 bp (Tabla 6). El tiempo de ejecución no aumenta con el tamaño de amplicón para la tecnología actual porque el tamaño de lectura es < 30-50 bases independientemente del tamaño del ácido nucleico diana a secuenciar. En algunas realizaciones, la tecnología proporcionada en este documento produce una secuencia de 2000 bp en un tiempo que es un orden de magnitud menor que la tecnología convencional (véase, p. ej., la Tabla 6). En algunas realizaciones, la tecnología proporcionada en este documento proporciona una lectura de secuencia más larga con el mismo tiempo de ejecución que la tecnología convencional.

Tabla 5 - Comparación para la secuenciación de amplicones largos de 1000 bp

	Biblioteca SOD ^a	Biblioteca Illumina TruSeq	Biblioteca de ADN _g de iones
Nº de muestras por ejecución	8	8	1
Nº de amplicones por muestra	50	50	50
Tamaño promedio de amplicones (pb)	1000	1000	1000
Longitud requerida de lectura de SBS	1 x 50	2 x 250 (extremo de par)	1 x 400 (bidireccional)
Tiempo total de ejecución (horas)	3	37	4 4
Media cobertura para cada amplicón por muestra	1974	---	---
Rendimiento (nº de muestras con 1000x de cobertura/hora)	5,3	---	---
Puntajes de calidad (porcentaje de lecturas con puntaje > Q30)	> 90%	---	---
a) La biblioteca SOD se ejecuta en un MiSeq con el kit de reactivos de secuenciación v2			

Tabla 6 - Comparación para la secuenciación de amplicones largos de 2000 bp

Aplicación de lectura larga (tamaño de inserción de 2Kb)			
	Biblioteca SOD ^a	Illumina Lib	Ion PGM Lib
5	Nº de muestras por ejecución	8	4
	Nº de amplicones por muestra	50	50
	Tamaño promedio de amplicones (bp)	2000	2000
10	Longitud requerida de lectura de SBS	1 x 50	1 x 400
	Tiempo total de ejecución (horas)	3 en MiSeq; 0,5 en PGM	7,25
15	Media cobertura para cada amplicón por muestra	962	---
	Rendimiento (nº de muestras con cobertura 2000x/hora)	2,6	---
	Puntajes de calidad (porcentaje de lecturas con puntaje >Q30)	> 90%	---
20	Coste por ejecución (reactivo seq y chip solo, \$)	725	---
	Coste por muestra (\$)	90,63	---
25	a) El tiempo de preparación de la biblioteca SOD para un tamaño de inserción más largo es más largo en algunas realizaciones (p. ej., de < 6,5 horas a < 8,5 horas) b) Tecnología Illumina "Moleculo"		

Ejemplo 4 - Verificación del concepto de datos obtenidos usando una biblioteca de modelos

30 **[0188]** Durante el desarrollo de las realizaciones de la tecnología proporcionada aquí, se recopilaron datos para verificar la tecnología usando una biblioteca de modelos. Como se muestra en la Figura 4, se construye una secuencia de consenso de ~ 127 bp a partir de una colección de lecturas de ~ 35 bp producidas de acuerdo con las realizaciones de la tecnología proporcionada. El tiempo de ejecución de secuenciación calculado en un aparato de secuenciación de ADN Illumina MiSeq para producir la secuencia de ~ 127 bp usando una biblioteca producida por la tecnología proporcionada en este documento es de aproximadamente 2,5 horas. Usando la tecnología convencional para proporcionar la biblioteca, un tiempo de ejecución de ~ 13 horas produce la misma lectura de secuencia de < 127 bp.

Ejemplo 5 - Generación de escalera usando terminación dNTP de 3'-O-propargilo

40 **[0189]** Durante el desarrollo de formas de realización de la tecnología proporcionada en el presente documento, se llevaron a cabo experimentos para evaluar la generación de fragmentos de ácido nucleico terminados en una reacción que comprende una mezcla de 3'-O-propargilo-dNTP y dNTP naturales (estándar). En particular, se llevaron a cabo experimentos para evaluar la generación de fragmentos terminados en cada posición dentro de la región diana mediante la incorporación de 3'-O-propargilo-dNTP de terminación de cadena por la polimerasa de ADN durante la síntesis. Los ensayos de extensión de polimerasa se realizaron usando un ácido nucleico de plantilla que tiene una secuencia de KRAS humano (p. ej., KRAS exón 2 y secuencias intrónicas flanqueantes) y un cebador complementario:

Plantilla KRAS Exón 2 (SEQ ID NO: 1)

50 TTATTATAAGGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAA
GAGTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGGTAA
ATCTTGTTTTAATATGCATATTACTGGTGCAGGACCATTCT

55 R_ke2 trP1_T_bio (SEQ ID NO: 2)

bTAAUCCTCTCTATGGGCAGTCGGTGATAGAATGGTCCTGCACCAGTAA

60 **[0190]** En la secuencia de cebador R_ke2_trP1_T_bio (SEQ ID NO: 2), una "b" indica una modificación de biotina y una "U" indica una modificación de desoxiuridina. La incorporación de los cebadores en productos de extensión produce productos de extensión que comprenden un uracilo. El uracilo es útil, p. ej., para la escisión del producto (p. ej., usando reactivos de escisión de uracilo) en varias manipulaciones biológicas moleculares (p. ej., escisión del producto de un soporte sólido).

65 **[0191]** Los experimentos se realizaron usando una mezcla de dNTP naturales y los cuatro 3'-O-propargilo-dNTP en

una sola reacción. La mezcla de reacción de generación de fragmentos de ADN comprendía 20 mM Tris-HCl, 10 mM (NH₄)SO₄, 10 mM KCl, 2 mM MnCl₂, 0,1% Triton X-100, 1000 pmol dATP, 1000 pmol dCTP, 1000 pmol dGTP, 1000 pmol de dTTP, 100 pmol de 3'-O-propargilo-dATP, 100 pmol de 3'-O-propargilo-dCTP, 100 pmol de 3'-O-propargilo-dGTP, 100 pmol de 3'-O-propargilo dTTP, 6,25 pmol de cebador R_ke2_trP1_T_bio (SEQ ID NO: 2) y 2 unidades de polimerasa de ADN THERMINATOR II (New England BioLabs) en un volumen de reacción de 25 µl. Se usaron 0,5 pmoles de amplicón purificado correspondiente a una región en el exón 2 de KRAS (SEQ ID NO: 1) como plantilla. La reacción de extensión de polimerasa se termocicló calentando a 95°C durante 2 minutos, seguido de 45 ciclos a 95°C durante 15 segundos, 55°C durante 25 segundos y 65°C durante 35 segundos.

[0192] Después de la reacción de extensión de polimerasa, se usó 1 µl de la mezcla de reacción directamente para el análisis del tamaño del fragmento de ADN usando electroforesis en gel (Bioanalizador Agilent 2100 y Chip de ensayo de ADN de alta sensibilidad). El análisis del tamaño de los fragmentos de los productos de reacción indicó que la reacción de generación de fragmentos produjo con éxito una escalera de fragmentos de ácido nucleico con los tamaños esperados.

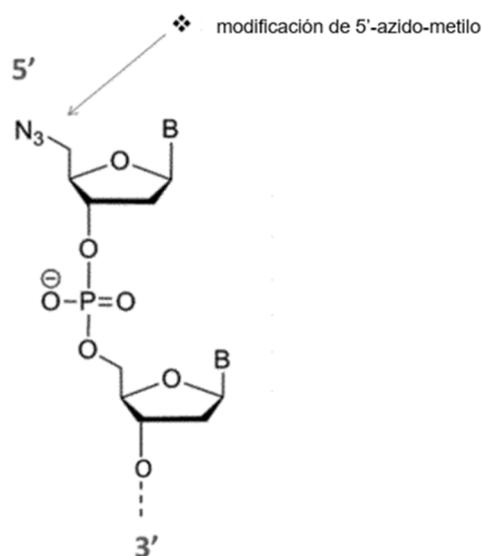
Ejemplo 6 - Síntesis de oligonucleótido modificado por 5'-azido-metilo

[0193] Durante el desarrollo de formas de realización de la tecnología proporcionada en el presente documento, un oligonucleótido que comprende una modificación de 5'-azido-metilo fue sintetizado y caracterizado. La síntesis del oligonucleótido modificado se realizó usando síntesis química de fosforamidita. En la última etapa sintética, se usó síntesis química de fosforamidita para incorporar una fosforamidita 5'-yodo-dT en la posición 5' terminal. El oligonucleótido unido al soporte sólido en la columna de reacción se trató entonces como sigue.

[0194] En primer lugar, azida de sodio (30 mg) se resuspendió en DMF seco (1 ml), se calentó durante 3 h a 55°C, y se enfrió a temperatura ambiente. El sobrenadante se recogió con una jeringa de 1 ml y se pasó de un lado a otro a través de la columna de reacción que comprende el oligonucleótido modificado con yodo 5' y se incubó durante la noche a temperatura ambiente (ambiente). Después de la incubación, la columna se lavó con DMF seco, se lavó con acetonitrilo y luego se secó con gas argón. El oligonucleótido modificado con 5'-azido-metilo resultante se escindió del soporte sólido y se desprotegió calentando en amoniaco acuoso durante 5 horas a 55°C. El producto final era un oligonucleótido que tenía la secuencia que se muestra a continuación:



[0195] El "Az" indica la modificación de azido-metilo en el extremo 5' (p. ej., modificación 5'-azido-metilo), p. ej., para proporcionar un oligonucleótido que tiene una estructura de acuerdo con



donde B es la base del nucleótido (p. ej., adenina, guanina, timina, citosina o una nucleobase natural o sintética, p. ej., una purina modificada tal como hipoxantina, xantina, 7-metilguanina; una pirimidina modificada tal como 5,6-dihidouracilo, 5-metilcitosina, 5-hidroximetilcitosina, etc.).

Ejemplo 7 - Conjugación del oligonucleótido modificado por 5'-azido-metilo y fragmentos de de ácido nucleico modificados por 3'-O-propargilo

[0196] Durante el desarrollo de formas de realización de la tecnología proporcionada en el presente documento, se realizaron experimentos para probar la conjugación de un oligonucleótido modificado con 5'-azido-metilo (p. ej., véase el Ejemplo 6) a fragmentos de ácido nucleico modificado con 3'-O-propargilo (p. ej., véase el Ejemplo 5) mediante química de clic. En particular, se llevaron a cabo experimentos en los que un oligonucleótido modificado con 5'-azido-metilo se conjugó químicamente con fragmentos de ADN modificado con 3'-O-propargilo usando química de cicloadición de 1,3-dipolar alquino-azida catalizada por cobre ("química de clic").

[0197] La química de clic se realizó usando reactivos disponibles comercialmente (baseclick GmbH, kit de recarga Oligo-Click-M) de acuerdo con las instrucciones del fabricante. Brevemente, aproximadamente 0,1 pmol de fragmentos de ADN modificados con 3'-O-propargilo que comprenden una modificación de 5'-biotina se hicieron reaccionar con aproximadamente 500 pmol de oligonucleótido modificado con 5'-azido-metilo usando el reactivo químico de clic en un volumen total de 10 µL. La mezcla de reacción se incubó a 45°C durante 30 minutos. Después de la incubación, el sobrenadante se transfirió a un nuevo tubo de microcentrífuga y se añadió un volumen de 40 µL del tampón de unión y lavado suministrado comercialmente (p. ej., 1 M NaCl, Tris-HCl 10 mM, EDTA 1 mM, pH 7,5). El producto de reacción conjugado se aisló del exceso de oligonucleótido modificado con 5'-azido-metilo incubando la mezcla de reacción química de clic con perlas magnéticas recubiertas con estreptavidina (Dynabeads, MyOne Streptavidin C1, Life Technologies) a temperatura ambiente (ambiente) durante 15 minutos. Las perlas se separaron del sobrenadante usando un imán y se retiró el sobrenadante. Posteriormente, las perlas se lavaron dos veces usando el tampón de unión y lavado y luego se resuspendieron en 25 µL de tampón TE (Tris-HCl 10 mM, EDTA 0,1 mM, pH aproximadamente 8).

[0198] El producto se escindió del soporte sólido (perla) usando la escisión de uracilo (glucosilasa de uracilo y endonucleasa VIII, enzimática). En particular, se usaron reactivos de escisión de uracilo para escindir los productos de reacción en el sitio de la modificación de desoxiuridina ubicada cerca de la ubicación 5'-terminal del producto conjugado (ver SEQ ID NO: 2-5). Finalmente, el sobrenadante que comprende el producto conjugado se purificó usando Ampure XP (Beckman Coulter) siguiendo el protocolo del fabricante y se eluyó en 20 µL de tampón TE.

Ejemplo 8 - Amplificación de producto conjugado

[0199] Durante el desarrollo de formas de realización de la tecnología descrita en el presente documento, se realizaron experimentos para caracterizar la conjugación química del oligonucleótido modificado por 5'-azido-metilo a los fragmentos de ácido nucleico modificados por 3'-O-propargilo y para evaluar el enlace de triazol como un imitador de un enlace fosfodiéster natural en una cadena principal de ácido nucleico. Para probar la capacidad de una polimerasa para reconocer el producto conjugado como plantilla y atravesar el enlace triazol durante la síntesis, los cebadores de PCR se diseñaron para producir amplicones que abarcan el enlace triazol de los productos de conjugación:

Cebador 1 CCTCTCTATGGGCAGTCGGTGAT SEQ ID NO: 4

Cebador 2 CCATCTCATCCCTGCGTGTCTC SEQ ID NO: 5

[0200] Se usó una premezcla de PCR disponible comercialmente (KAPA 2G HS, KAPA Biosystems) para proporcionar una mezcla de reacción de 25 µl que comprende, además de los componentes proporcionados por la mezcla (p. ej., tampón, polimerasa, dNTP), 0,25 µM de Cebador 1 (SEQ ID NO: 4), 0,25 µM de Cebador 2 (SEQ ID NO: 5) y 2 µL de producto conjugado purificado (ver Ejemplo 7) como plantilla para la amplificación. La mezcla de reacción se sometió a un ciclo térmico incubando la muestra a 95°C durante 5 minutos, seguido de 30 ciclos de 98°C durante 20 segundos, 60°C durante 30 segundos y 72°C durante 20 segundos. Los productos de amplificación se analizaron mediante electroforesis en gel (p. ej., usando un sistema Agilent Bioanalyzer 2100 y un chip de ADN de alta sensibilidad) para determinar las distribuciones de tamaño de los productos de reacción.

[0201] El análisis de los productos de amplificación indicó que la reacción de amplificación produjo exitosamente amplicones usando los productos conjugados de la reacción química de clic (ver Ejemplo 7) como plantillas para la amplificación. En particular, el análisis de los productos de amplificación indicó que la polimerasa procesada a lo largo de la plantilla y a través del enlace triazol para producir amplicones a partir de la plantilla. Además, la amplificación produjo una población heterogénea de amplicones que tenían un rango de tamaños correspondiente a los tamaños esperados producidos por la amplificación de los fragmentos de ADN terminados específicos de la base mediante la incorporación del 3'-O-propargilo-dNTP. El análisis de fragmentos también mostró el aumento de tamaño de fragmento apropiado correspondiente a treinta y una (31) bases adicionales del oligonucleótido conjugado con 5'-azido-metilo modificado.

Ejemplo 9: Ligadura de adaptadores NGS a productos de escalera de fragmentos

[0202] Durante el desarrollo de realizaciones de la tecnología proporcionada en este documento, se realizaron experimentos para secuenciar fragmentos de escalera producidos de acuerdo con la tecnología proporcionada en este documento (véase la Figura 5). Como paso inicial en la secuenciación, se realizaron experimentos para preparar una biblioteca de secuenciación usando productos de escalera de ADN generados en el Ejemplo 8 como entrada y un kit comercial para la preparación de muestras. Las bibliotecas de secuenciación se prepararon usando un kit de

preparación de muestras de ADN TRUSEQ NANO (Illumina, Inc.) siguiendo el protocolo del fabricante con la siguiente modificación. Después del paso de ligadura del adaptador, se realizaron dos rondas (en lugar de una ronda) de purificación basada en perlas utilizando una relación 1:1 (v/v) de muestra a mezcla de perlas. Se realizaron 8 ciclos de amplificación utilizando los cebadores de PCR Illumina proporcionados para enriquecer los productos ligados al adaptador siguiendo el protocolo del fabricante. La biblioteca de secuenciación final se analizó mediante electroforesis en gel (Bioanalizador Agilent 2100 y Chip de ensayo de ADN de alta sensibilidad). El análisis del tamaño del fragmento confirmó la generación exitosa de una biblioteca NGS (p. ej., para la secuenciación de Illumina) usando los productos de escalera de fragmentos del Ejemplo 8. Los datos indicaron que la biblioteca NGS tenía el aumento de tamaño de fragmento apropiado correspondiente a la adición de la Illumina de 126 bp adaptadores y, por lo tanto, que los adaptadores se ligaron adecuadamente a la escalera de fragmentos. La Figura 5 muestra un esquema de fragmentos de la biblioteca de secuenciación. En particular, los fragmentos comprenden un adaptador Illumina en ambos extremos, una o más secuencias universales y una secuencia diana.

Ejemplo 10 – Secuenciación

[0203] Durante el desarrollo de realizaciones de la tecnología proporcionada en el presente documento, se realizaron experimentos para secuenciar una biblioteca NGS ligada a adaptador, p. ej., una biblioteca de secuenciación preparada como se describe en el Ejemplo 9. La biblioteca producida según el Ejemplo 9 fue secuenciado con éxito usando un secuenciador Illumina MiSeq usando un kit de secuenciación por síntesis de 2 x 75 bp. El kit proporciona cebadores de secuencia complementarios a las secuencias adaptadoras. Después de la secuenciación, más del 89% de las lecturas tuvieron un puntaje de calidad de secuencia de Q30 o mejor.

[0204] Los datos recopilados de los experimentos indicaron que la población de fragmentos proporciona la alineación inequívoca de las lecturas de secuenciación cortas (30-50 bp) producidas por la tecnología. En particular, los fragmentos de ácido nucleico superpuestos proporcionaron lecturas que se alinearon y ensamblaron con éxito a pesar de su pequeño tamaño.

[0205] Los datos de secuencia fueron extraídos de la salida secuenciadora utilizando un conjunto de datos personalizados de procesamiento de flujo de trabajo que se acomoda para el diseño particular de la escalera de fragmento producida según la tecnología. Por ejemplo, el software personalizado identificó lecturas y lecturas procesadas para usar porciones de 40 bp de las 2 x lecturas de secuencia de 75 bp para la alineación de secuencia posterior. Los componentes particulares del software personalizado concatenan lecturas (p. ej., Archivos Read1 y Read2 FASTQ) producidos desde el secuenciador NGS; identificar la secuencia que se origina en la secuencia diana, la secuencia universal y los adaptadores (p. ej., identificar la secuencia que se origina en el 5'-azido-metiloligonucleótido); establecer un límite de extracción de secuencia usando el reconocimiento de patrones; extraer la secuencia diana de las lecturas de secuencia producidas por el secuenciador NGS; y alinear las secuencias (ver Figura 5).

Ejemplo 11 - Alineación de secuencia

[0206] Durante el desarrollo de las realizaciones de la tecnología proporcionada en el presente documento, se realizaron experimentos para alinear los datos de secuencia producidos a partir de una biblioteca NGS como se describe en este documento, producir una secuencia de consenso a partir de la alineación y alinear la secuencia de consenso con una secuencia de referencia. En particular, las lecturas de secuencia de 40 bp que se extrajeron de la salida de secuenciación de MiSeq se alinearon contra una secuencia de referencia (p. ej., una secuencia de 177 bp que comprende secuencias de intrón flanqueante parcial del exón 2 del gen KRAS humano).

[0207] La alineación de las lecturas de secuenciación de 40 bp se realizó usando CLC Genomics Workbench v7 con penalizaciones estrictas por desajustes e indels; los requisitos de coincidencia de longitud y similitud se establecieron adecuadamente de acuerdo con las instrucciones adjuntas para lecturas de 40 bp. Los resultados de alineación (Figura 6A) indicaron que las lecturas de secuencia de 40 bp proporcionaron una cobertura completa de toda la secuencia de referencia (177 bp). Además, la gráfica de profundidad de cobertura versus posición de secuencia tenía el perfil de cobertura "trapezoidal" esperado que se dilucidaba durante la simulación de alineación teórica (Figura 6B).

[0208] Estos resultados indican que una secuencia de secuenciación relativamente corta (p. ej., MiSeq con 30 a 50 ciclos de secuenciación por síntesis) produce una secuencia completa de alta calidad de la diana. Además, con ajustes a los métodos existentes, p. ej., diseñando cebadores para que se unan inmediatamente adyacentes al sitio diana, se puede maximizar la longitud de la secuencia de alta calidad. Además, la longitud de la secuencia de alta calidad también se puede maximizar con la generación apropiada de la escala de fragmentos para cubrir la longitud completa de la longitud total de la diana (p. ej., ajustando la proporción de 3'-O-propargilo-dNTPs a dNTPs; ver Ejemplo 12). En este ejemplo, 40 ciclos de secuenciación (para obtener 40 bases de secuencia) en el MiSeq tomaron aproximadamente 2,5 horas. Sin embargo, es importante destacar que la tecnología proporciona una mejora sobre las tecnologías existentes en el sentido de que el tiempo de ejecución del secuenciador no cambia según el tamaño de destino.

Ejemplo 12: Secuenciación y análisis de bibliotecas NGS

[0209] Durante el desarrollo de realizaciones de la tecnología proporcionada en el presente documento, se realizaron experimentos para controlar la distribución de tamaños de fragmentos de ácido nucleico terminados producidos en una reacción que comprende una mezcla de 3'-O-propargilo-dNTPs y dNTP naturales (estándar) ajustando la relación de 3'-O-propargilo-dNTP a dNTP naturales (estándar). Se contempló que la relación molar de 3'-O-propargilo-dNTPs y dNTPs naturales afecta la distribución del tamaño del fragmento debido a la competencia entre los 3'-O-propargilo-dNTPs (que terminan la extensión) y los dNTPs naturales (que alargan el producto de polimerasa) para su incorporación en el ácido nucleico sintetizado por la polimerasa.

[0210] En consecuencia, se realizaron experimentos en los que los productos de las reacciones de generación de escalera de fragmentos se evaluaron a diversas relaciones molares de 3'-O-propargilo-dNTP a dNTP naturales. Las reacciones de generación de escalera de fragmentos se realizaron usando relaciones molares 2:1, 10:1 y 100:1 de dNTP naturales a 3'-O-propargilo-dNTP. Las mezclas de reacción de generación de fragmentos utilizadas en estos experimentos comprendían 20 mM Tris-HCl, 10 mM (NH₄)SO₄, 10 mM KCl, 2 mM MnCl₂, 0,1% de Triton X-100, 1.000 pmol dATP, 1000 pmol dCTP, 1000 pmol dGTP, 1000 pmol dTTP, 6,25 pmol de cebador, 2 unidades de polimerasa de ADN Terminator II (New England BioLabs) y 0,5 pmol de amplicón purificado correspondiente a una región en el exón 2 de KRAS (SEQ ID NO: 1) como plantilla en un 25 µl de volumen de reacción final.

[0211] Además, las reacciones que probaron una proporción 2:1 de dNTPs naturales a 3'-O-propargilo-dNTPs comprendieron 500 pmol de 3'-O-propargilo-dATP, 500 pmol de 3'-O-propargilo-dCTP, 500 pmol de 3'-O-propargilo-dGTP, y 500 pmol de 3'-O-propargilo dTTP. Las reacciones que probaron una relación 10:1 de dNTPs naturales a 3'-O-propargilo-dNTPs comprendieron 100 pmol de 3'-O-propargilo dATP, 100 pmol de 3'-O-propargilo-dCTP, 100 pmol de 3'-O-propargilo-dGTP y 100 pmol de 3'-O-propargilo-dTTP. Las reacciones que probaron una proporción de 100:1 de dNTPs naturales a 3'-O-propargilo-dNTPs comprendieron 10 pmol de 3'-O-propargilo-dATP, 10 pmol de 3'-O-propargilo-dCTP, 10 pmol de 3'-O-propargilo-dGTP, y 10 pmol de 3'-O-propargilo-dTTP.

[0212] Las reacciones de extensión de polimerasa se sometieron a ciclos de temperatura mediante incubación a 95°C durante 2 minutos, seguido de 45 ciclos a 95°C durante 15 segundos, 55°C durante 25 segundos y 65°C durante 35 segundos. Después de la reacción de extensión de la polimerasa, los oligonucleótidos modificados con 5'-azido-metilo se conjugaron químicamente con los fragmentos de ácido nucleico terminados con 3'-O-propargilo-dN usando química de clic como se describe en el Ejemplo 6 y el Ejemplo 7. Después de la conjugación, los productos de conjugación se usaron como plantillas para la amplificación para producir amplicones correspondientes a los productos conjugados como se describe en el Ejemplo 8. El análisis del tamaño de fragmento se realizó en los productos conjugados.

[0213] El análisis del tamaño del fragmento de los productos de conjugación amplificados producidos a partir de los productos de las tres condiciones diferentes de relación molar indicó que el tamaño del fragmento dependía de la relación de 3'-O-propargilo-dNTPs a dNTPs naturales. El análisis de los tamaños de los fragmentos muestra un cambio en la distribución del tamaño de los fragmentos en función de las relaciones molares de dNTP a 3'-O-propargilo-dNTP. En la relación molar 2:1, se detectaron poblaciones más grandes de fragmentos más cortos en comparación con las otras dos condiciones de relación molar. En la relación molar 10:1, estaba presente una fracción mayor de fragmentos más largos en relación con la relación molar 2:1. En la relación molar 100:1, la población principal de fragmentos comprendía fragmentos de ADN más largos en relación con las otras dos relaciones molares.

[0214] Los fragmentos de escalera producidos con las tres relaciones molares diferentes se usaron como entradas separadas para generar bibliotecas NGS (Illumina) para la secuenciación en el secuenciador MiSeq como se describe en el Ejemplo 9. Además, las lecturas de secuencia se obtuvieron como se describe en el Ejemplo 10 y se extrajeron datos de la secuencia de la secuencia diana y analizaron como se describe en el Ejemplo 11.

[0215] Los perfiles de cobertura de las tres bibliotecas que se prepararon usando las tres relaciones molares diferentes de dNTP a 3'-O-propargilo-dNTP (relaciones molares de 2:1, 10:1 y 100:1) se correlacionaron con la distribución del tamaño del fragmento de escalera de ADN creada por las respectivas relaciones molares. Por ejemplo, se esperaba que la relación molar 2:1 de dNTP a 3'-O-propargilo-dNTP terminara la extensión de la polimerasa a una frecuencia alta debido a la abundancia relativamente alta de 3'-O-propargilo-dNTP y, por lo tanto, produce ácido nucleico fragmentos de escalera que son relativamente más cortos que a relaciones más altas de dNTP a 3'-O-propargilo-dNTP. Por el contrario, se esperaba que la relación molar 100:1 terminara la extensión de la polimerasa a una frecuencia baja debido a la abundancia relativamente baja de 3'-O-propargilo-dNTP y, por lo tanto, produce fragmentos de escalera de ácido nucleico que son relativamente más largos que a relaciones más bajas de dNTP a 3'-O-propargilo-dNTP.

[0216] Los datos recogidos del análisis de tamaño de fragmento de los productos de escalera de ADN generados usando las tres relaciones molares diferentes confirmaron estas predicciones. En particular, los datos indican que variar la relación molar de dNTP a 3'-O-propargilo-dNTP proporciona el control del tamaño del fragmento de escalera de ADN.

[0217] Además, la secuenciación de los productos de escalera de ADN generados usando las tres relaciones molares diferentes y el análisis de la secuencia producida a partir de los productos de escalera mostró que los perfiles de cobertura de secuencia se correlacionaron con la relación molar de dNTP a 3'-O-propargilo-dNTP usado durante la

generación de la escalera de ADN. En particular, los datos indicaron que la relación molar 2:1 proporcionó más cobertura de la secuencia cerca del sitio de unión del cebador de secuenciación y la relación molar 100:1 proporcionó más cobertura adicional del sitio de unión del cebador de secuenciación. En consecuencia, la tecnología proporciona la capacidad de controlar la generación de fragmentos de escalera de ADN para una variedad de aplicaciones de secuenciación. En particular, aumentar la cobertura distante del sitio de unión del cebador de secuenciación es útil para la secuenciación de aplicaciones relacionadas con aplicaciones de secuenciación largas (p. ej., más de 100 pares de bases). La secuenciación usando múltiples bibliotecas de secuenciación producidas a diferentes proporciones molares proporciona datos de secuencia que tienen una alta cobertura de secuencias que están cerca, intermedia y lejos del sitio de unión del cebador de secuenciación.

Ejemplo 13: Etiquetado con cebadores que comprenden una secuencia de índice

[0218] Durante el desarrollo de realizaciones de la tecnología proporcionada en el presente documento, se realizaron experimentos para evaluar el uso de secuencias de índice o código de barras para rastrear y construir la secuencia de la plantilla diana original a partir de la secuencia producida a partir de la generación de bibliotecas, NGS y alineación. En el primer conjunto de experimentos, los ácidos nucleicos diana se copiaron y se marcaron mediante reacciones de extensión de polimerasa usando cebadores específicos de la diana que comprenden una secuencia de índice de identificación única. Como se usa en este documento, este y otros enfoques de códigos de barras moleculares similares se denominan "reacción de copiar y etiquetar" o "reacción de copiar y etiquetar ID".

[0219] En este esquema, se diseñó un cebador de extensión de polimerasa que comprende dos regiones (Figura 7): una región 3' que comprende una secuencia de cebado específica de la diana y una región 5' que comprende dos secuencias universales diferentes (p. ej., la secuencia universal A y secuencia universal B) que flanquea una secuencia degenerada (p. ej., que comprende 8 bp). Los cebadores de oligonucleótidos se sintetizaron de acuerdo con este esquema y se usaron en reacciones de extensión de polimerasa con un segundo oligonucleótido diseñado para detener la extensión de polimerasa y, por lo tanto, "copiar y etiquetar" solo la región de interés:

Cebador de extensión de polimerasa Eg_e19_R_SOD_v03-01-bio (SEQ ID NO: 6)

bTAAUTAGTGGCTGACGGGTATCTCTCACCTTTNNNNNNNCAGACATGAGAAAAGGTGGGC

Extensión de la polimerasa bloqueador Eg_e19_SOD_SC-200_v1 (SEQ ID NO: 7)

C*A*ATTGTGAGATGGTGCCACATGCTGCam

[0220] En las secuencias del cebador de extensión de la polimerasa y la extensión de la polimerasa bloqueador que se usa en la reacción de extensión de polimerasa durante el procedimiento de "copia y etiqueta" (SEQ ID NOs: 6 y 7 anteriores), una "b" indica una modificación de 5'-biotina, una "U" indica una modificación de desoxiuridina, un "*" indica un enlace de fosforotioato y "am" indica una modificación de 3'-amino.

[0221] Las reacciones de extensión de polimerasa se realizaron usando un kit de mezcla maestra de polimerasa de alta fidelidad disponible en el mercado (kit KAPA HiFi HotStart PCR, KAPA Biosystems) para producir una mezcla de reacción que comprende 1 pmol de cebador de extensión de polimerasa (p. ej., Eg_e19_R_SOD_v03-01-bio), 1 pmol de bloqueador de la extensión de polimerasa (p. ej., Eg_e19_SOD_SC-200_v1), y 100 ng de ADN genómico purificado extraído de una línea celular de adenocarcinoma de pulmón humano/cáncer de pulmón de células no pequeñas (línea celular NCI-H1975 disponible de ATCC bajo acceso CRL-5908) en un volumen de reacción de 25 µl. Las reacciones de extensión de polimerasa se incubaron a 95°C durante 2 minutos, 98°C durante 30 segundos, 58°C durante 90 segundos y 65°C durante 30 segundos. La dNTP y la polimerasa KAPA HiFi se agregaron inmediatamente después de completar la etapa de incubación a 58°C.

[0222] Los productos de reacción de extensión de polimerasa se purificaron usando purificación basada en perlas (Ampure XP, Beckman Coulter) siguiendo el protocolo del fabricante para eliminar cebadores de extensión de polimerasa, bloqueadores de extensión de polimerasa y otros componentes de reacción de extensión. A continuación, se ha usado una purificación basada en la captura de fase sólida utilizando microesferas magnéticas recubiertas con estreptavidina (Dynabeads, MyOne Estreptavidina C1, Life Technologies) para aislar los productos de reacción de extensión de polimerasa de la plantilla de ADN genómico. Después de aislar los productos de reacción de extensión de polimerasa, se añadió un tampón de unión y lavado 2 x (2 M NaCl, Tris-HCl 20 mM, EDTA 2 mM, pH 7,5) al eluyente de la purificación de perlas a 1:1 (v/v) cociente e incubado con las cuentas de estreptavidina a temperatura ambiente (ambiente) durante 15 minutos. Las perlas se separaron del sobrenadante usando un imán y se retiró el sobrenadante. A continuación, las perlas se lavaron dos veces usando tampón de unión y lavado y se resuspendieron en 25 µl de tampón TE (Tris-HCl 10 mM, EDTA 0,1 mM, pH aproximadamente 8). Las perlas se incubaron con una solución de NaOH 0,1 M y NaCl 0,1 M durante 1 minuto para eliminar cualquier rastro de ADN genómico restante. Luego se separaron las perlas del sobrenadante usando un imán (se descartó el sobrenadante), se lavaron dos veces usando tampón de unión y lavado, y se resuspendieron en 25 µl de tampón TE (Tris-HCl 10 mM, EDTA 0,1 mM, pH aproximadamente 8).

[0223] Finalmente, para liberar el producto unido a la perla, se usó un sistema de escisión de uracilo (Uracil Glycosylase and Endonuclease VIII, Enzymatics) para escindir el producto de extensión de la polimerasa unida a la perla en la modificación de desoxiuridina incorporada en el extremo 5' del producto de extensión de polimerasa como resultado del cebador de la extensión de polimerasa (véase la SEQ ID NO: 6). El sobrenadante que comprende el producto de extensión de polimerasa se purificó usando Ampure XP (Beckman Coulter) siguiendo el protocolo del fabricante y se eluyó en 20 µl de tampón TE.

[0224] Los cebadores de amplificación Uni_R_v2 y e19_F_v1 se diseñaron, sintetizaron y usaron para amplificar el producto de extensión de polimerasa purificado para confirmar la generación del producto de copia y etiqueta como se describe esquemáticamente en la Figura 8. Los cebadores de amplificación Uni_R_v2 y SC-240_COM_v1 se usaron para confirmar que el bloqueador de extensión de polimerasa bloqueó efectivamente la extensión de polimerasa más allá del sitio en donde el bloqueador de extensión de polimerasa se une a la plantilla.

Uni_R_v2 (SEQ ID NO: 8)
AGTGGCTGACGGGTATCTCTC

e19_F_v1 (SEQ ID NO: 9)
TGCCAGTTAACGTCTTCCTTC

SC-240_COM_v1 (SEQ ID NO: 10)
ATCACTGGGCAGCATGTGG

[0225] Se realizaron dos reacciones de amplificación en el producto. Una primera reacción comprendió los cebadores Uni_R_v2 y e19_F_v1, que amplifican tanto los productos de extensión de polimerasa bloqueados (a través del bloqueador de la polimerasa) como los no bloqueados. Una segunda reacción comprendió los cebadores Uni_R_v2 y SC-240_COM_v1, que amplifican solo el producto de extensión de polimerasa no bloqueado. Los dos tipos de mezclas de reacción se produjeron usando una mezcla de amplificación disponible comercialmente (KAPA 2G HS, KAPA Biosystems) y 0,25 µM de cada cebador (como se indicó anteriormente para las dos reacciones) en un volumen de reacción final de 25 µl. Se usó un volumen de 5 µl de producto de extensión de polimerasa purificado como plantilla para cada reacción de amplificación. Las reacciones de amplificación se termociclaron incubando las mezclas de reacción a 95°C durante 5 minutos, seguido de 30 ciclos de 98°C durante 20 segundos, 60°C durante 30 segundos y 72°C durante 20 segundos. Los productos de amplificación se analizaron mediante electroforesis en gel (p. ej., usando un sistema Agilent Bioanalyzer 2100 y un chip de ADN de alta sensibilidad) para determinar las distribuciones de tamaño de fragmento.

[0226] Los datos recogidos del análisis de tamaño de fragmento indicaron que la reacción de amplificación que comprende los cebadores Uni_R_v2 y e19_F_v1 produjo un producto del tamaño esperado. Además, los datos también indicaron que la reacción de amplificación que comprende los cebadores Uni_R_v2 y SC-240_COM_v1 no generó un producto detectable, lo que indica que el bloqueador de la extensión de polimerasa detiene efectivamente la reacción de polimerasa. En consecuencia, la tecnología proporciona un control preciso de la reacción de copia y etiqueta para producir productos solo de una región diana de interés.

Ejemplo 14: Etiquetado con adaptadores que comprenden una secuencia de índice

[0227] Además, en un segundo conjunto de experimentos realizados durante el desarrollo de las realizaciones descritas en el presente documento, los ácidos nucleicos diana se copiaron y posteriormente se marcaron mediante ligadura de adaptador usando adaptadores que comprenden una secuencia de índice de identificación única. En este esquema de código de barras molecular basado en la ligadura del adaptador (véase, p. ej., la Figura 9), se construyó un adaptador de ADN usando dos oligonucleótidos. El primer oligonucleótido se diseñó para tener un tramo de secuencia degenerada (p. ej., que comprende de 8 a 12 bases) flanqueado tanto en el extremo 5' como en el extremo 3' por dos secuencias universales diferentes (p. ej., la secuencia universal A y la secuencia universal B; ver Figura 9). El segundo oligonucleótido se diseñó para comprender una secuencia universal C (p. ej., en el extremo 5') y una secuencia (p. ej., en el extremo 3') que es complementaria a la secuencia universal B y que tiene una T adicional en el extremo 3' - posición terminal. Para producir el adaptador de ADN, los dos oligonucleótidos se mezclaron en cantidades molares iguales, se incubaron a 95°C durante 5 minutos y luego se enfriaron lentamente a temperatura ambiente (ambiente) para proporcionar una hibridación eficiente de las porciones complementarias de los dos oligonucleótidos (p. ej., secuencia universal B y su secuencia complementaria). La ligadura de estos adaptadores al ADN diana proporciona el "etiquetado ID" único de cada molécula de ADN diana individual (p. ej., cada amplicón de PCR individual), p. ej., en una reacción que comprende un exceso molar de adaptadores de secuencia de etiqueta ID únicos en relación con el número de moléculas diana individuales.

[0228] Se llevaron a cabo experimentos para probar las realizaciones de esta tecnología usando los siguientes oligonucleótidos:

ST-adN10-phos-v1 (SEQ ID NO: 11)

pGTGGCTGACGGGTATCTCTC>NNNNNNNNNNNATCACCGACTGCCCATAGAGAGG

ST-ad-T-v1 (SEQ ID NO: 12)

GCACTGGATCACGTCATACCTACGAGAGATACCCGTCAGCCA*C*T

5 **[0229]** En las secuencias de los dos oligonucleótidos usados para formar el adaptador (SEQ ID NOs: 11 y 12 anteriores), una "p" indica una modificación 5'-fosfato, una "N" indica una posición de base degenerada (p. ej., la posición puede ser A, C, G o T), y un "*" indica un enlace de fosforotioato.

10 **[0230]** Como primer paso, se realizó una reacción de amplificación para amplificar una región de 158 bp en el exón 18 (con secuencia de intrón flanqueante) del gen EGFR humano usando los siguientes cebadores:

E_e18_f_v1p (SEQ ID NO: 13)

pCCAGTGGAGAAGCTCCCAAC

15 **E_e18_r_v1p (SEQ ID NO: 14)**
pCAGACCATGAGAGGCCCTG

20 **[0231]** En las secuencias de los dos cebadores EGFR (SEQ ID NO: 13 y 14 anteriores), una "p" indica una modificación 5'-fosfato. Las mezclas de reacción se produjeron usando un kit de mezcla maestra de PCR disponible comercialmente (kit KAPA 2G HotStart PCR, KAPA Biosystems), 10 pmol de cada uno de los cebadores EGFR (SEQ ID NO: 13 y 14) y 10 ng de ADN genómico purificado extraído de una línea celular de adenocarcinoma de pulmón humano/cáncer de pulmón de células no pequeñas (línea celular NCI-H1975 disponible de ATCC con el número de acceso CRL-5908) en un volumen de reacción de 25 µl. Las mezclas de reacción se termociclaron incubando a 95°C durante 2 minutos, seguido de 23 ciclos de 98°C durante 20 segundos, 63°C durante 30 segundos y 68°C durante 20 segundos. Después de la amplificación, se usó 1 µl de la mezcla de reacción directamente para el análisis del tamaño del fragmento de ADN usando electroforesis en gel (p. ej., Agilent 2100 Bioanalyzer y Chip de ensayo de ADN de alta sensibilidad). Los datos recopilados del análisis de fragmentos indicaron que la amplificación generó un producto que tenía el tamaño esperado de 158 bp.

30 **[0232]** A continuación, el producto de amplificación se purificó para eliminar cebadores no incorporados y componentes de reacción de amplificación usando un método de purificación basado en perlas (Ampure XP, Beckman Coulter) siguiendo el protocolo del fabricante.

35 **[0233]** Después de la purificación, un adaptador que comprende una secuencia de índice (p. ej., como se describió anteriormente) se ligó al amplicón. El amplicón producido por la reacción de amplificación anterior comprendía un fosfato 5' (p. ej., de la incorporación de los cebadores modificados 5'-fosfato) y un saliente 3'-dA (p. ej., a partir de una polimerasa de ADN que añade una A no moldeada en el extremo 3' de los productos de extensión). La reacción de ligadura se realizó usando un kit de ligadura disponible comercialmente (T4 DNA Ligase-Rapid, Enzymatics). En particular, se produjo una mezcla de reacción de ligadura utilizando el kit de tampón de ligadura "Rapid", 25 pmoles de adaptador y aproximadamente 0,25 pmoles del amplicón en un volumen de reacción de 50 µl.

45 **[0234]** Después de la reacción de ligadura, la mezcla de reacción de ligadura se incubó a 25°C durante 10 minutos y se purificó inmediatamente dos veces usando purificación basada en perlas (Ampure XP, Beckman Coulter) siguiendo el protocolo del fabricante, excepto que el volumen de entrada de la muestra a la solución de perlas el volumen se cambió de 1:1,8 a 1:1.

50 **[0235]** El producto ligado purificado se usó como plantilla en una amplificación de enriquecimiento de ciclo limitado (p. ej., 8 ciclos) para amplificar el producto ligado (Figura 10). La reacción de amplificación comprendió cebadores diseñados para amplificar el producto ligado que comprende la porción de etiqueta "etiqueta ID" (p. ej., 10 bases degeneradas) y que tiene una longitud esperada de 249 bp:

PCR1 (SEQ ID NO: 15)

CCTCTCTATGGGCAGTCCGGTGAT

55 **ST-PCR1-R-v1 (SEQ ID NO: 16)**
GCACTGGATCACGTCATACCTAC

60 **[0236]** La amplificación se realizó usando un kit de mezcla maestra de PCR de polimerasa de alta fidelidad disponible en el mercado (KAPA HiFi HotStart PCR kit, KAPA Biosystems) para producir una mezcla de reacción que comprende 0,25 µM de cada cebador y el producto ligado al adaptador purificado como plantilla en un volumen de reacción de 25 µl. Las mezclas de reacción de amplificación se termociclaron incubando a 95°C durante 5 minutos, seguido de 8 ciclos de 98°C durante 20 segundos, 60°C durante 30 segundos y 72°C durante 20 segundos. Después de la amplificación, se usó 1 µl de la mezcla de reacción directamente para el análisis del tamaño del fragmento por electroforesis en gel (Bioanalyzer Agilent 2100 y chip de ensayo de ADN de alta sensibilidad). Datos recopilados del análisis de fragmento indicó que la amplificación produjo un amplicón del tamaño esperado del producto ligado al

adaptador (p. ej., un amplicón de 249 bp que comprende una porción correspondiente al amplicón EGFR de 158 bp producido anteriormente y un adaptador ligado).

Ejemplo 15 - Circularización del ácido nucleico diana

[0237] Durante el desarrollo de las realizaciones de la tecnología proporcionada en este documento, se realizaron experimentos para evaluar una técnica molecular basada en la ligadura intramolecular (circularización) del ácido nucleico diana para orientar diferentes regiones del ácido nucleico diana en una disposición específica: el método comprende la circularización de un ácido nucleico diana, que coloca una secuencia conocida (p. ej., una secuencia de cebado universal) adyacente a una secuencia desconocida (p. ej., una región de interés para consultar, p. ej., por secuenciación) en orientación específica (Figura 11).

[0238] En estos experimentos, las reacciones de circularización se realizaron usando un kit de ligasa de ADNss disponible en el mercado (CircLigase II, Epicenter-Illumina) siguiendo el protocolo del fabricante. Los experimentos probaron plantillas de entrada sintéticas que eran oligonucleótidos ("ultramers") que tenían longitudes de 100, 150 y 200 bases:

Ultramer-200bp (SEQ ID NO: 17)

pGCAGCATGTGGCACCATCTCACAATTGCCAGTTAACGTCTTCCTTCTCTCTGGTGAGAAAGTTAAAA
 TTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGATGTG
 AGTTTCTGCTTTGCTGTGTGGGGTCCATGGCTCTGAACCTCAGGCCACCTTTTCTCATGTCTG

Ultramer-150bp (SEQ ID NO: 18)

pGCAGCATGTGGCACCATCTCACAATTGCCAGTTAACGTCTTCCTTCTCTCTATCTCCGAAAGCCAAC
 AAGGAAATCCTCGATGTGAGTTTCTGCTTTGCTGTGTGGGGTCCATGGCTCTGAACCTCAGGCCAC
 CTTTTCTCATGTCTG

Ultramer-100bp (SEQ ID NO: 19)

pGCAGCATGTGGCACCATCTCACAATTGCCAGTTAACGTCTTCCTTCTCTCTGATGTGAGTTTCTGCT
 TTGCTTCCTCAGGCCACCTTTTCTCATGTCTG

En las secuencias de los ultrameros (SEQ ID NOs: 17, 18 y 19 anteriores), una "p" indica una modificación 5'-fosfato.

[0239] Después de la reacción de circularización, los productos se trataron con exonucleasa I y III (NEB) durante 30 minutos a 37°C para eliminar la plantilla no circularizada. Después del tratamiento con exonucleasa, las exonucleasas se inactivaron incubando a 80°C durante 10 minutos. Para confirmar la circularización de las plantillas, los cebadores se diseñaron para amplificar productos de amplificación específicos del círculo (Figura 12):

e19_F_v1 (SEQ ID NO: 20)

TGCCAGTTAACGTCTTCCTTCT

e19_circ_v1 (SEQ ID NO: 21)

G*A*TGGTGCCACATGCTGC

[0240] En las secuencias de los cebadores de plantilla circular (SEQ ID NOs: 20 y 21 anteriores), un "*" indica un enlace de fosforotioato.

[0241] Mezclas de reacción de amplificación se producen utilizando Taq-Gold (Abbott Molecular), 0,2 µM de cada cebador, y uno de los tres productos de reacción de diferentes tamaños como plantilla en 25 µL de volumen de reacción. Las mezclas de reacción se termociclaron incubando a 95°C durante 5 minutos, seguido de 38 ciclos de 98°C durante 20 segundos, 60°C durante 30 segundos y 68°C durante 30 segundos. Después de la amplificación, se usaron 10 µL de la mezcla de reacción directamente para el análisis del tamaño del fragmento de ADN mediante electroforesis en gel usando geles de agarosa prefabricados al 2% (E-Gel EX 2% Agarose Gel, Life Technologies). Los datos recopilados indicaron que la amplificación produjo un producto del tamaño esperado a partir de las plantillas circulares, confirmando así la generación de ácidos nucleicos circulares a partir de los tres ultrameros de prueba. Además, la ausencia de productos específicos de círculo en controles negativos que comprenden plantillas lineales indica que los cebadores producen productos específicos de círculo.

[0242] Aunque la tecnología se ha descrito en relación con realizaciones ejemplares específicas, debe entenderse que la invención como se reivindica no debe limitarse indebidamente a tales realizaciones específicas.

REIVINDICACIONES

1. Un método para determinar una secuencia de nucleótidos diana, comprendiendo el método:

5 a) generar una biblioteca de secuenciación de próxima generación mediante:

1) amplificar una secuencia de nucleótidos diana usando un cebador que comprende una secuencia específica de diana y una secuencia universal A para proporcionar un amplicón, en donde el amplicón puede ser monocatenario o bicatenario; y

10 2) ligar un primer oligonucleótido adaptador que comprende una secuencia universal B al amplicón para formar un adaptador-amplicón; y 3) generar una biblioteca de fragmentos en escalera que comprende una pluralidad de fragmentos para usar como una biblioteca de secuenciación de próxima generación, en donde la biblioteca de fragmentos en escalera se genera usando un análogo de nucleótido 3'-O-alquilino; y

15 b) determinar una secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos escalera, comprendiendo dicha secuencia de nucleótidos una subsecuencia de nucleótidos de la secuencia de nucleótidos diana.

2. El método de la reivindicación 1, en donde el cebador comprende además una secuencia de nucleótidos de código de barras asociada con el ácido nucleico diana y el método comprende además determinar una secuencia de nucleótidos de código de barras del fragmento de la biblioteca de fragmentos de escalera, y opcionalmente en donde la secuencia de nucleótidos de código de barras comprende 1 a 20 nucleótidos.

3. El método de la reivindicación 2, en donde determinar la secuencia de nucleótidos del código de barras del fragmento de la biblioteca de fragmentos de escalera comprende usar un cebador oligonucleotídico complementario a la secuencia universal B.

4. El método de la reivindicación 1, que comprende además ligar un segundo oligonucleótido adaptador que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar una biblioteca de secuenciación de próxima generación y en donde determinar la secuencia de nucleótidos de un fragmento de la biblioteca de fragmentos de escalera comprende usar un cebador oligonucleotídico complementario a la secuencia universal C, y que opcionalmente comprende uso de una reacción química de clic para ligar el segundo oligonucleótido adaptador que comprende una secuencia universal C a los extremos 3' de los fragmentos de la biblioteca de fragmentos de escalera para generar la biblioteca de secuenciación de próxima generación.

5. El método de la reivindicación 1, en donde:

(i) el primer oligonucleótido adaptador se liga al extremo 3' del amplicón;

(ii) el primer oligonucleótido adaptador comprende de 10 a 80 nucleótidos;

40 (iii) el análogo de nucleótido de 3'-O-alquilino es un análogo de nucleótido de 3'-O-propargilo;

(iv) generar la biblioteca de fragmentos de escalera comprende usar un cebador oligonucleotídico complementario a la secuencia universal A; y/o

(v) la secuencia de nucleótidos del fragmento de la biblioteca de fragmentos de escalera comprende de 15 a 1000 nucleótidos.

6. El método de la reivindicación 2 que comprende asociar la secuencia de nucleótidos del código de barras con una fuente de la secuencia de nucleótidos diana.

7. El método de la reivindicación 2, que comprende además secuencias de nucleótidos agrupadas de fragmentos de la biblioteca de fragmentos de escalera que tiene la misma secuencia de nucleótidos de código de barras.

8. El método de la reivindicación 1, que comprende además ensamblar una pluralidad de secuencias de nucleótidos de fragmentos de la biblioteca de fragmentos de escalera para proporcionar una secuencia de consenso.

9. El método de la reivindicación 8 que comprende además mapear la secuencia de consenso a una secuencia de referencia.

10. El método de la reivindicación 8, en donde la secuencia de consenso retiene la información de fase y/o enlace del ácido nucleico diana.

11. Una composición que comprende fragmentos de ADN superpuestos para usar como una biblioteca de secuenciación de próxima generación para obtener una secuencia de un ácido nucleico diana, comprendiendo la composición n ácidos nucleicos, en donde cada ácido nucleico comprende una subsecuencia de nucleótidos del ácido nucleico diana y en donde cada ácido nucleico comprende un análogo de nucleótido 3'-O-alquilino.

65

12. La composición de la reivindicación 11 en donde el análogo de nucleótido de 3'-O-alquínilo es análogo de nucleótido de 3'-O-propargilo.

5 13. La composición de la reivindicación 11, en donde un ácido nucleico comprende un adaptador unido a un análogo de nucleótido mediante un enlace formado por química de clic.

10

15

20

25

30

35

40

45

50

55

60

65

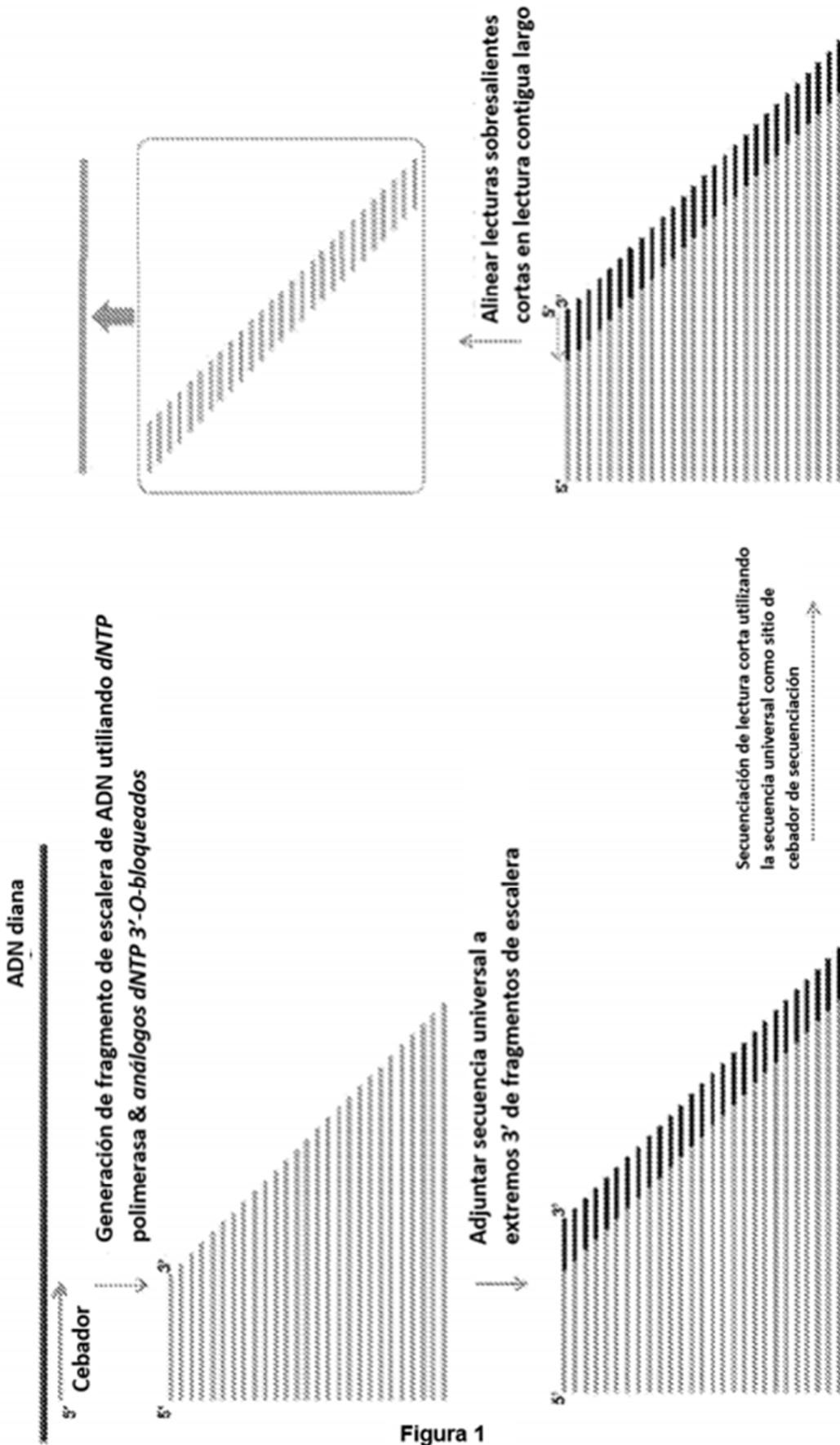


Figura 1

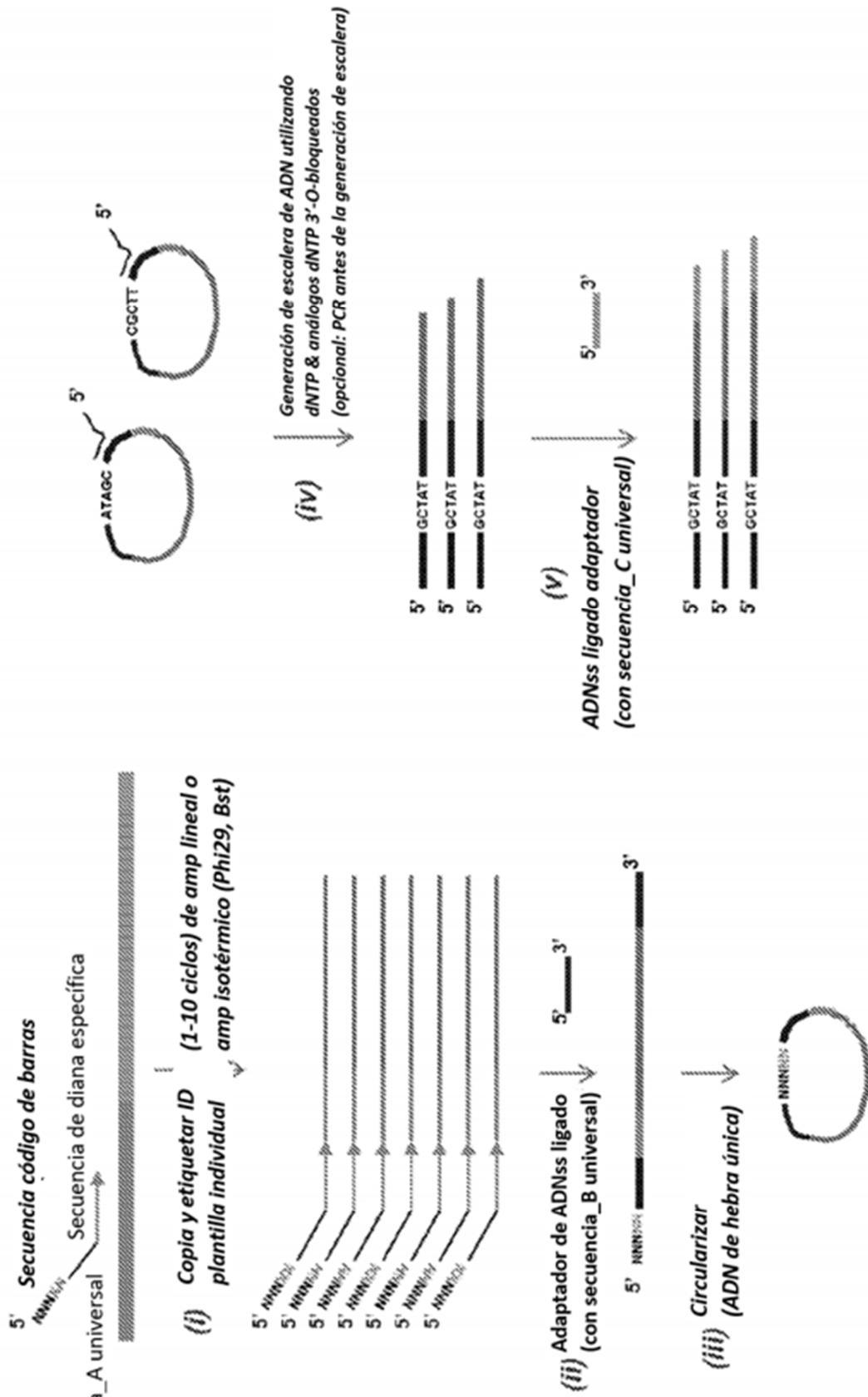


Figura 2A

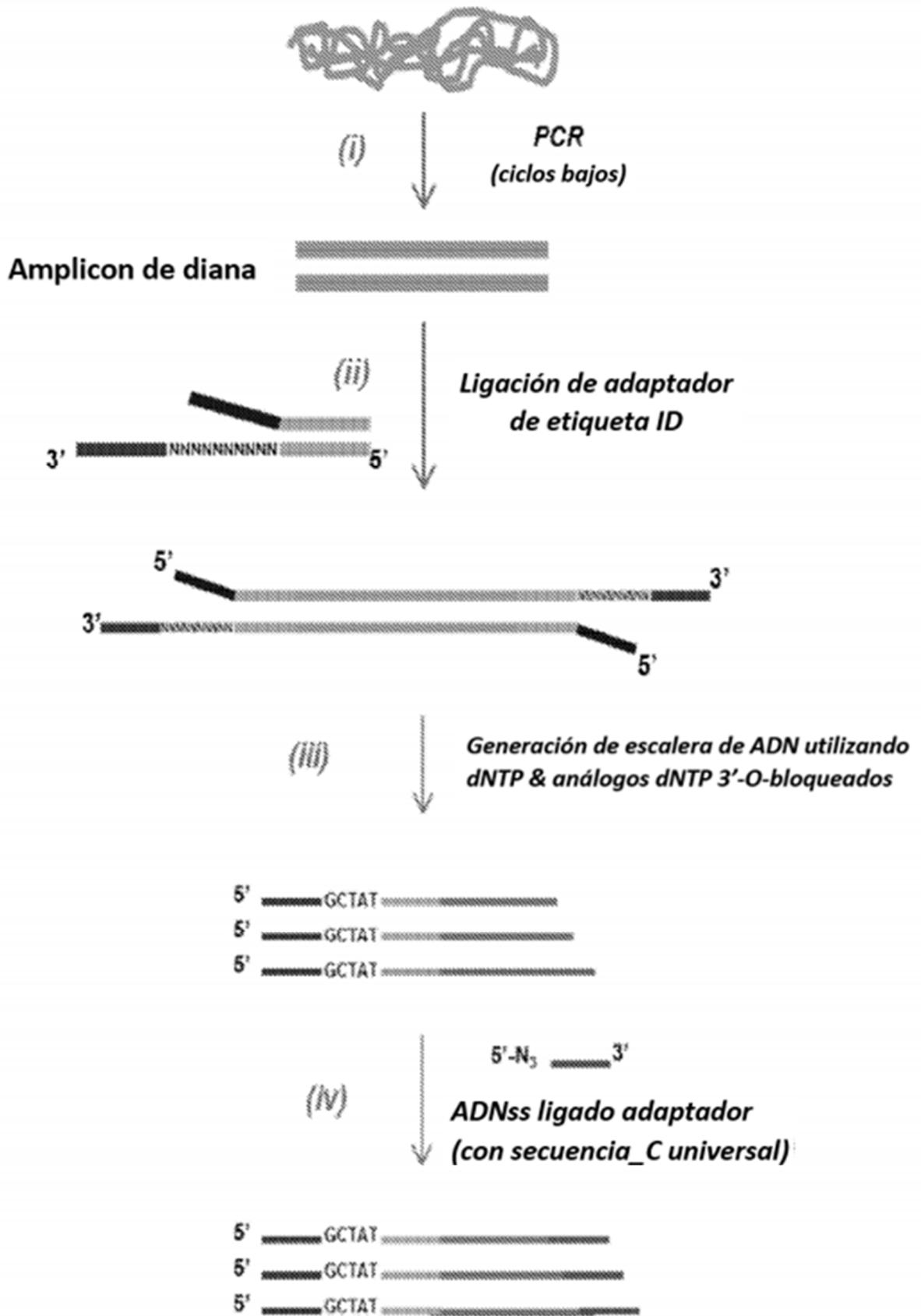


Figura 2C

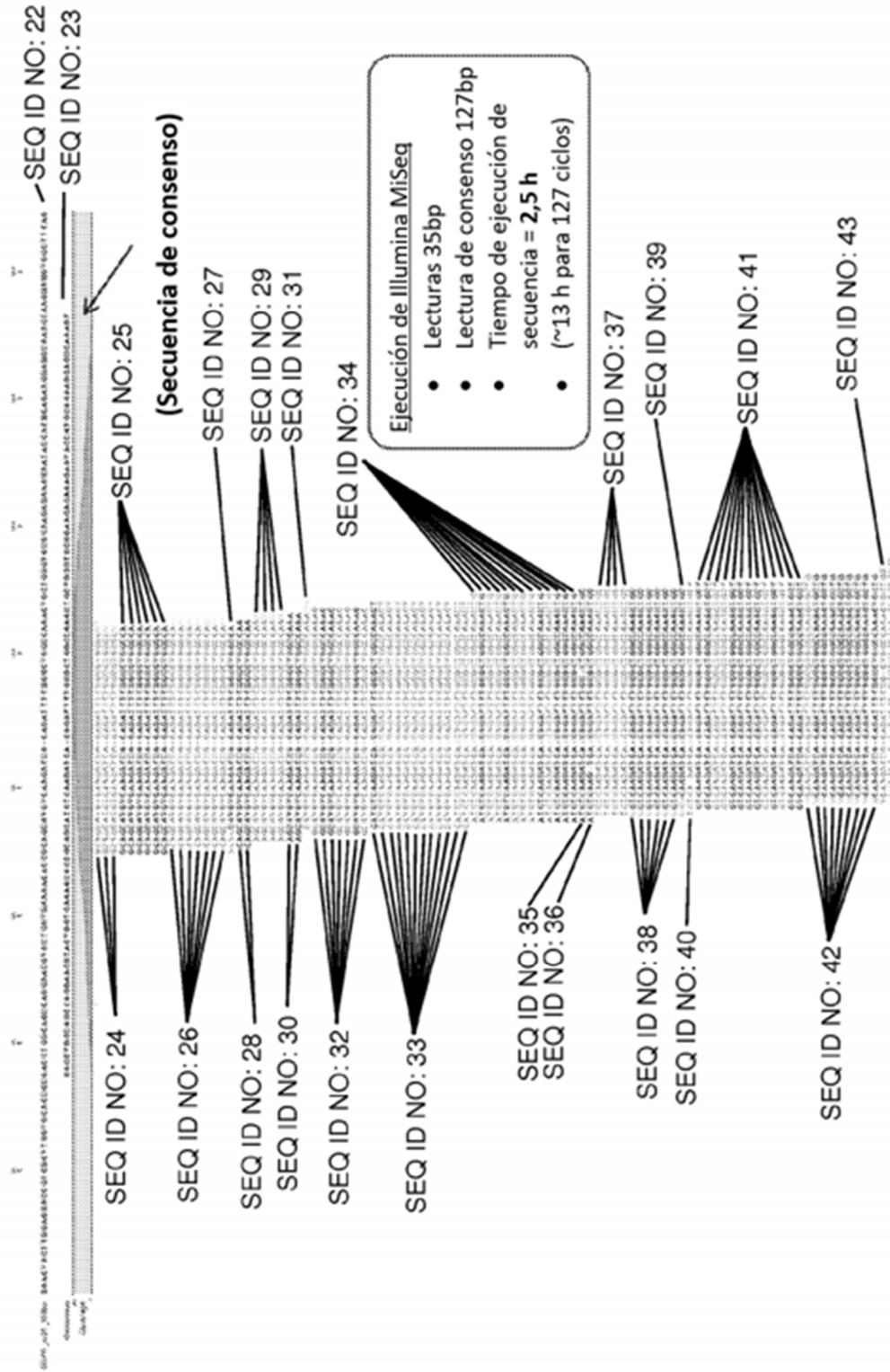


Figura 4

SEQ ID NO: 26

Biblioteca de NGS y esquema de secuenciación

Flujo de trabajo de extracción de secuencias

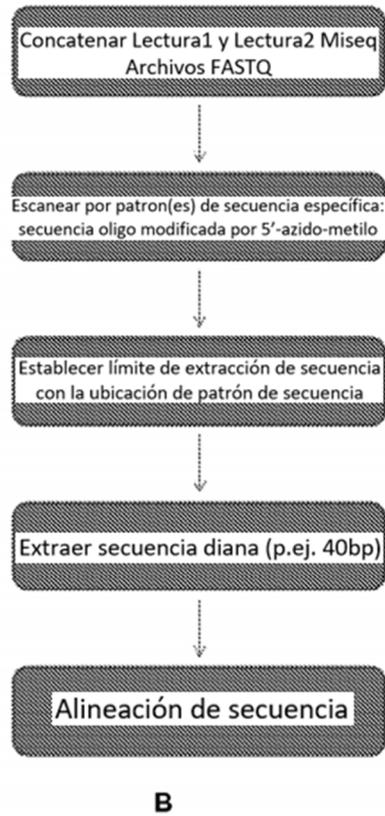
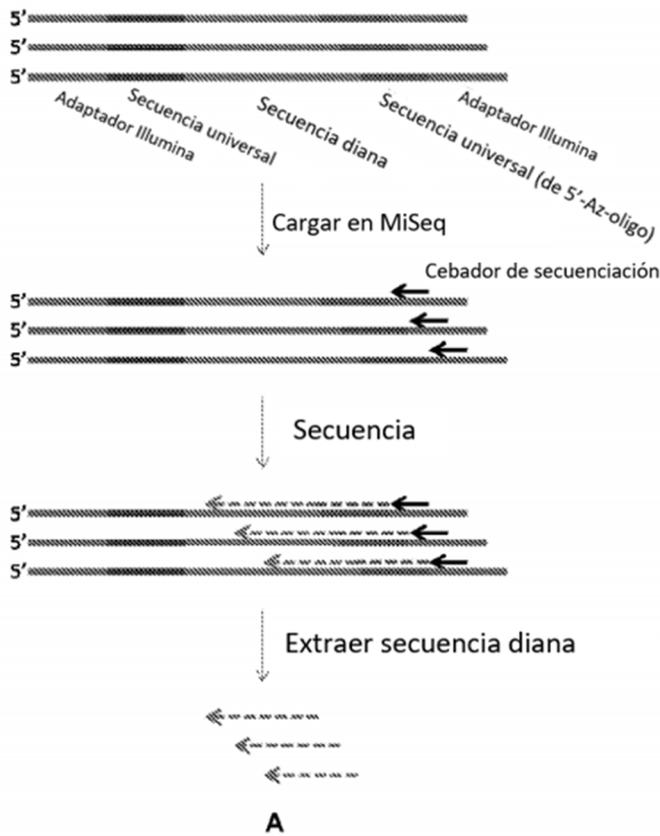
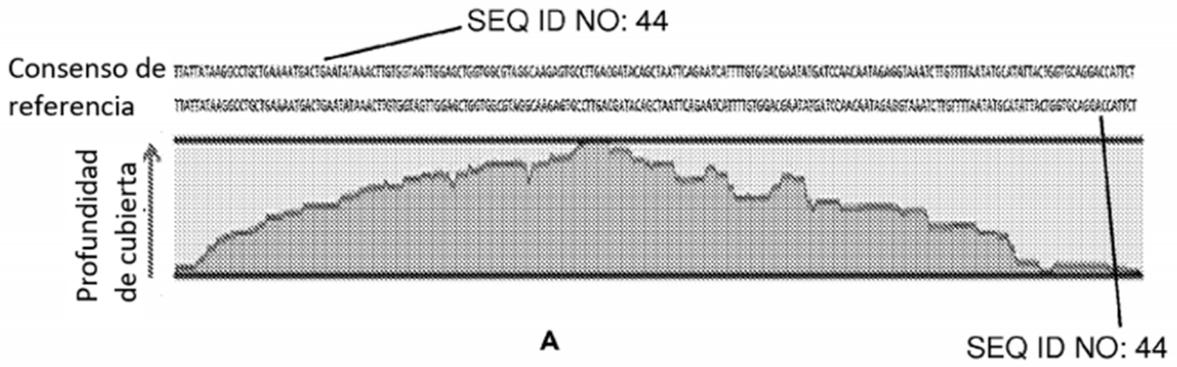


Figura 5



ACGGCTCGTTGGATGCTAGCTGATCGCGAA — SEQ ID NO: 45
 CGGCTCGCTGGATGCTAGCTGATCGCGAAT — SEQ ID NO: 46
 GGCTCGCTGGATGCTGGCTGATCGCGAATG — SEQ ID NO: 47
 GCACGCTGGATGCTAGCTGATCGCGAATGC — SEQ ID NO: 48
 CTCGCTGGATGCTAGCTGATCGAGAATGCA — SEQ ID NO: 49

 ACGGCTCGCTGGATGCTAGCTGATCGCGAATGCA — SEQ ID NO: 50

Secuencia de referencia

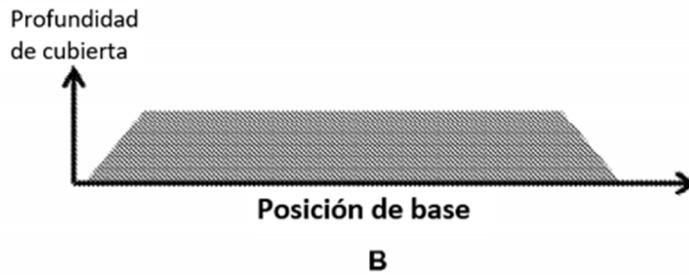


Figura 6

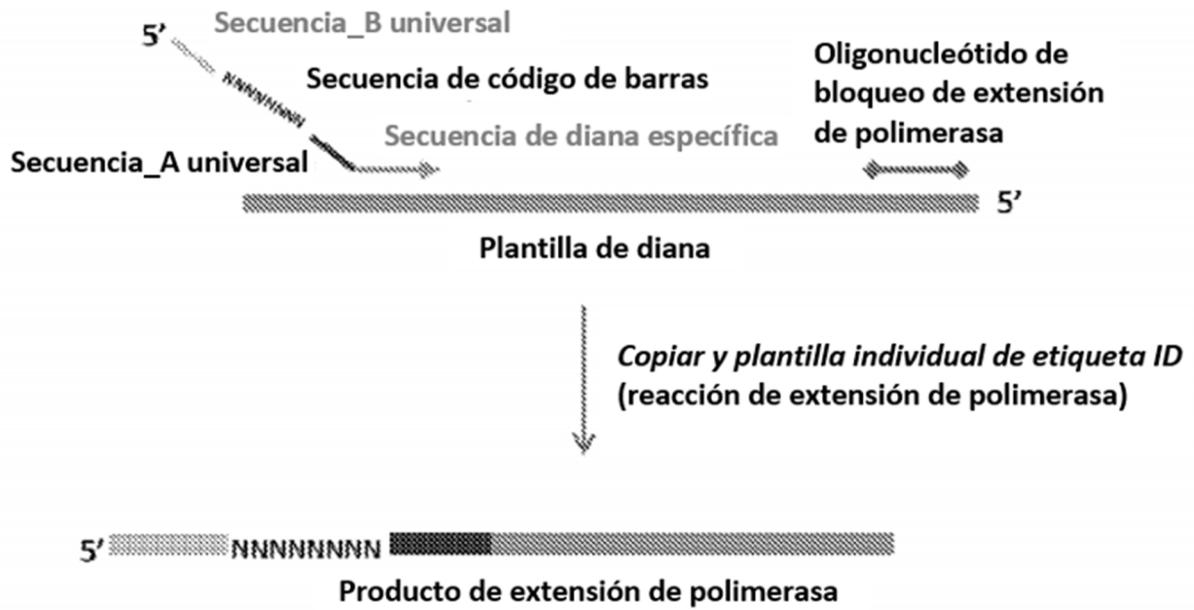
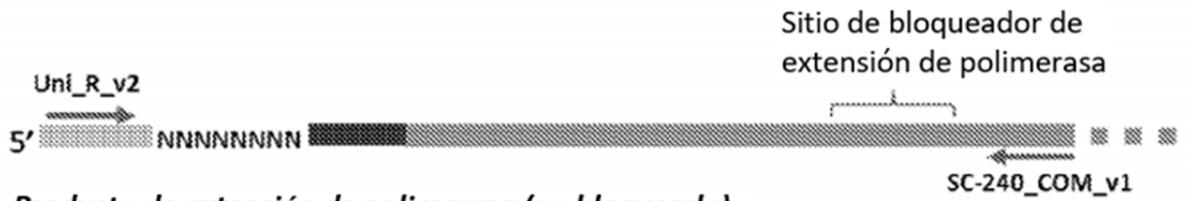


Figura 7

Detección de productos de extensión de polimerasa



Producto de extensión de polimerasa (bloqueado)



Producto de extensión de polimerasa (no bloqueado)

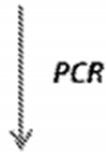


Figura 8

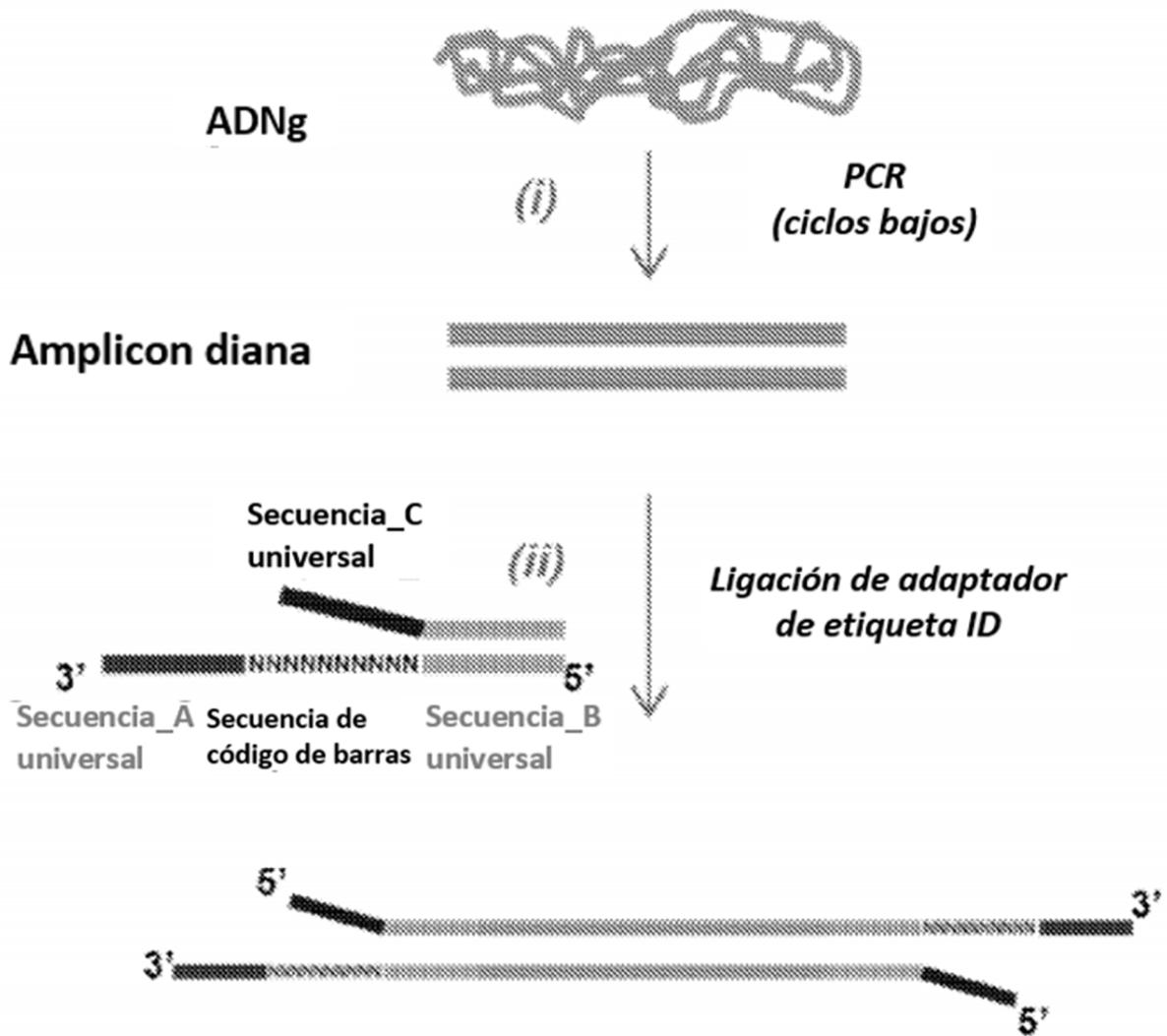
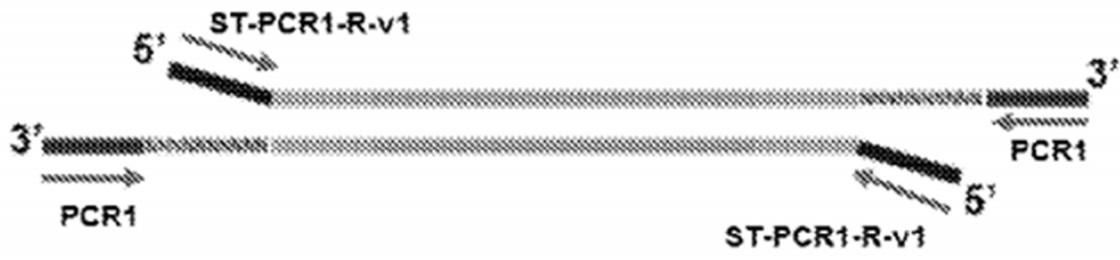


Figura 9



(Producto ligado a adaptador)

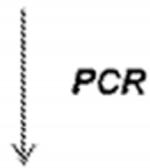
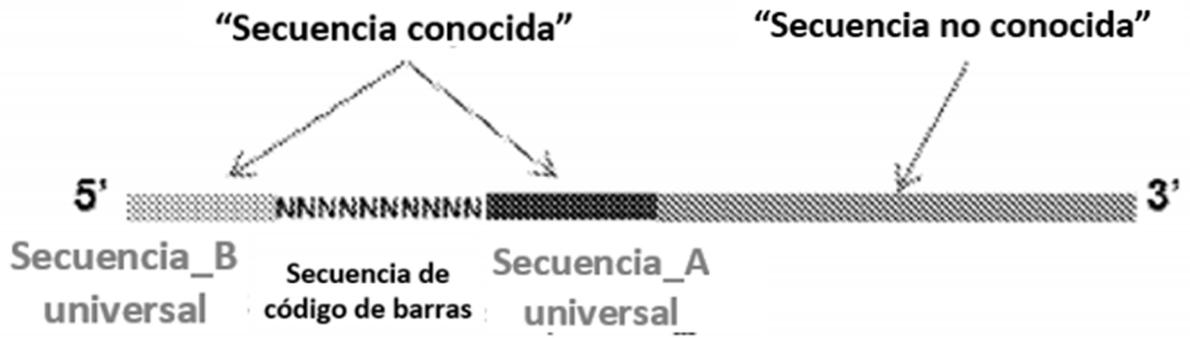


Figura 10



Ligación intramolecular de ADNss (circularización)



Figura 11

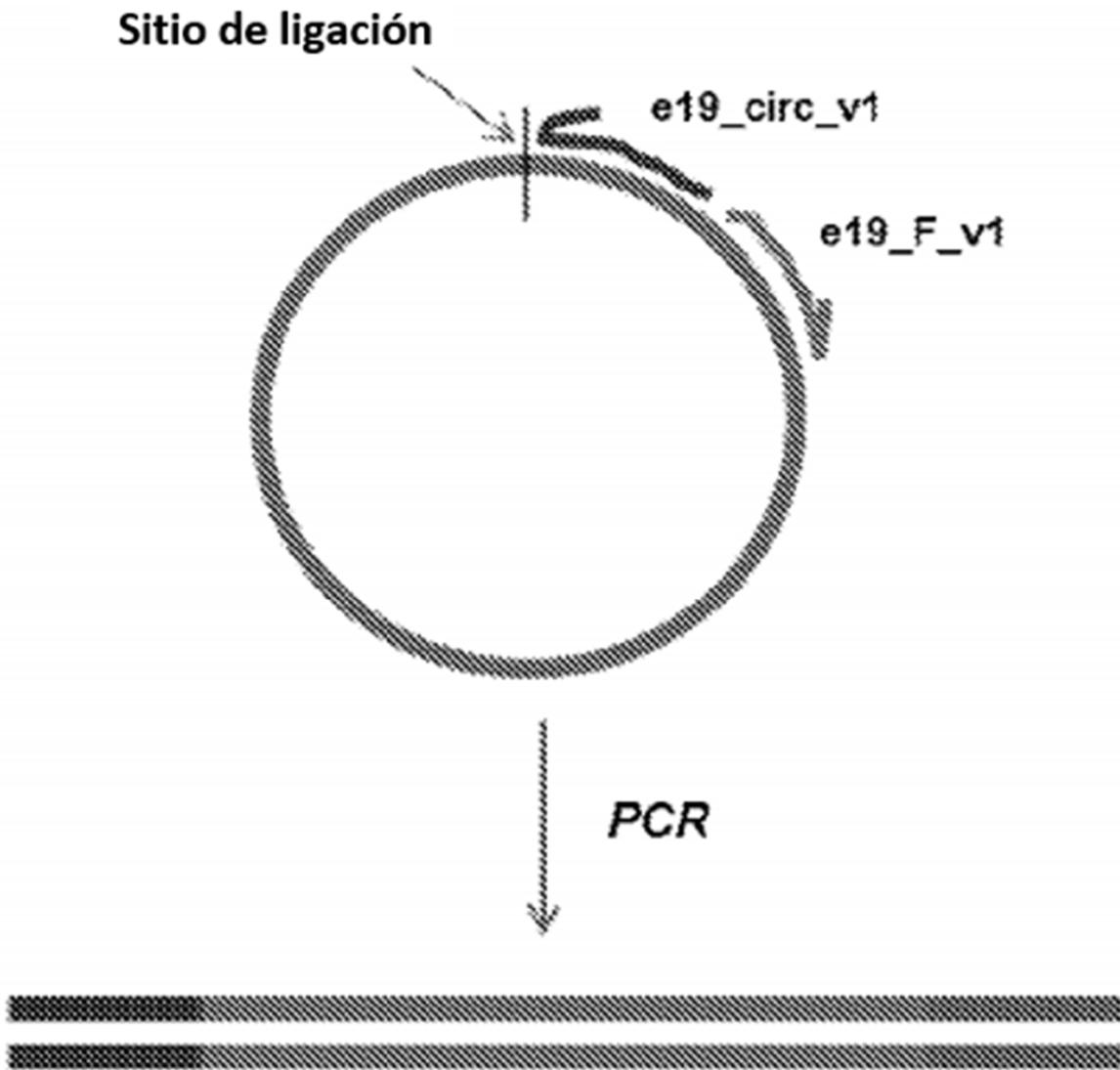


Figura 12