

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 811 518**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.06.2015 PCT/US2015/037653**

87 Fecha y número de publicación internacional: **30.12.2015 WO15200609**

96 Fecha de presentación y número de la solicitud europea: **25.06.2015 E 15744405 (0)**

97 Fecha y número de publicación de la concesión europea: **13.05.2020 EP 3161137**

54 Título: **Preparación de bibliotecas de ácido nucleico marcado usando protocolo de adición en un solo tubo**

30 Prioridad:

26.06.2014 US 201462017786 P

21.07.2014 US 201462027198 P

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

12.03.2021

73 Titular/es:

**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

72 Inventor/es:

**KAPER, FIONA y
CANN, GORDON**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 811 518 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

Preparación de bibliotecas de ácido nucleico marcado usando protocolo de adición en un solo tubo

La presente descripción se refiere generalmente a métodos para preparar una biblioteca de fragmentos de ácido nucleico, y más específicamente a métodos para preparar una biblioteca de fragmentos de ácido nucleico en un solo tubo usando proteasas, para una variedad de aplicaciones que incluyen, p. ej., secuenciación de ADN de próxima generación.

Antecedentes

Hay una variedad de métodos y aplicaciones para los cuales es deseable generar una biblioteca de ácido nucleico fragmentado y marcado, p. ej., para usar como plantillas en la secuenciación de ADN y/o para el análisis de la variación del número de copias.

Las tecnologías de secuenciación de ADN de "próxima generación" desarrolladas recientemente, tales como las desarrolladas por Illumina, Inc. (San Diego, CA), permiten generar datos de secuencias de hasta millones de plantillas de secuenciación en una sola secuencia usando un formato multiplexado o masivo paralelo. Esta naturaleza masiva paralela de la secuenciación de "próxima generación" requiere la generación de bibliotecas de fragmentos de ácido nucleico que contienen una colección o población de fragmentos de ácido nucleico a partir de una muestra de ácido nucleico objetivo, p. ej., un ADN genómico. Más importante aún, requiere que la combinación de estos fragmentos de ácido nucleico presente secuencias que sean cualitativa y/o cuantitativamente representativas de la secuencia de la muestra de ácido nucleico objetivo. Cuando la muestra de ácido nucleico proviene de células, los métodos actuales para generar una biblioteca de fragmentos de ácido nucleico típicamente requieren una etapa separada para aislar el ácido nucleico objetivo de las células, antes de la fragmentación del ácido nucleico. Esta etapa de extracción de ácido nucleico normalmente es un desperdicio de la muestra de ácido nucleico objetivo, y generalmente hace que el ácido nucleico preparado no pueda representar cualitativamente el ácido nucleico objetivo de la muestra. Esto se convierte en un problema particularmente importante cuando la cantidad de muestra es limitada o difícil de obtener. Para resolver este problema, algunos métodos actuales utilizan la amplificación de ácido nucleico antes de la fragmentación. Sin embargo, la amplificación no puede garantizar la representatividad del ácido nucleico objetivo ya que el ácido nucleico objetivo todavía se pierde parcialmente durante la extracción antes de la amplificación. El documento WO 2013/131962 describe un método de fragmentación.

Por lo tanto, existe la necesidad de nuevos métodos que permitan la preparación rápida y eficiente de la biblioteca de fragmentos de ácido nucleico. La presente descripción aborda esta necesidad proporcionando métodos para preparar una biblioteca de fragmentos de ácido nucleico en una única mezcla de reacción, p. ej., en un solo tubo, usando proteasas. Se proporcionan ventajas relacionadas también.

Resumen

En un aspecto, se proporciona en el presente documento un método para preparar una biblioteca de fragmentos de ácido nucleico marcados que incluye (a) poner en contacto una población de células directamente con un reactivo de lisis para generar un lisado celular, en donde el reactivo de lisis tiene una o más proteasas, y en donde el lisado celular contiene un ácido nucleico objetivo; (b) inactivar la una o más proteasas para formar un lisado celular inactivado, y (c) aplicar directamente al menos una transposasa y al menos una composición de extremos de transposón que contiene una cadena transferida al lisado celular inactivado en condiciones donde el ácido nucleico objetivo y la composición de extremos de transposón experimentan una reacción de transposición para generar una mezcla, en donde (i) el ácido nucleico objetivo se fragmenta para generar una pluralidad de fragmentos de ácido nucleico objetivo, (ii) la cadena transferida de la composición de extremos de transposón se une a los extremos 5' de cada una de una pluralidad de fragmentos de ácido nucleico objetivo para generar una pluralidad de fragmentos de ácido nucleico objetivo marcados en 5' y (iii) el ácido nucleico objetivo es un ADN bicatenario, y en donde el ácido nucleico objetivo sigue siendo el ADN bicatenario antes de aplicar una transposasa y una composición de extremos de transposón en la etapa (c).

En algunas realizaciones, las etapas (a), (b) y (c) proporcionados en el presente documento se realizan en una única mezcla de reacción, p. ej., en un tubo. En algunas realizaciones, la población de células es una población mínima de células. En algunas realizaciones, la población mínima de células contiene una, dos, tres, cuatro o cinco células.

En algunas realizaciones, la una o más proteasas se seleccionan de un grupo que consiste en serina proteasas, treonina proteasas, cisteína proteasas, aspartato proteasas, ácido glutámico proteasas y metaloproteasas. En algunas realizaciones, la una o más proteasas son subtilisinas y variantes de las mismas. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 0.1 mg/ml a 10 mg/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 0.1 mg/ml a 2.5 mg/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 0.5 mg/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 4.5 mUA/ml a 500 mUA/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 22.5 mUA/ml.

En algunas realizaciones, la población de células se pone en contacto con el reactivo de lisis de pH 7.0 a pH 10.0 en

la etapa (a). En algunas realizaciones, la población de células se pone en contacto con el reactivo de lisis a pH de 7.0 a pH de 9.0.

5 En algunas realizaciones, la una o más proteasas se inactivan aumentando la temperatura en la etapa (b). En algunas realizaciones, la una o más proteasas se inactivan aumentando la temperatura a 50°C-80°C. En algunas realizaciones, la una o más proteasas se inactivan aumentando la temperatura a 70°C. En algunas realizaciones, la una o más proteasas se inactivan añadiendo uno o más inhibidores de la una o más proteasas.

En algunas realizaciones, el reactivo de lisis incluye uno o más detergentes. En algunas realizaciones, el uno o más detergentes son detergentes no iónicos. En algunas realizaciones, el uno o más detergentes incluyen Triton.

10 El ácido nucleico objetivo es un ADN bicatenario, y en el que el ácido nucleico objetivo permanece como el ADN bicatenario antes de aplicar una composición de trasposasa y extremos de trasposón en la etapa (c). En algunas realizaciones, el ácido nucleico objetivo es ADN genómico. En algunas realizaciones, el ácido nucleico objetivo contiene ADN cromosómico o un fragmento del mismo. En algunas realizaciones, el ácido nucleico objetivo incluye un genoma o un genoma parcial.

15 En algunas realizaciones, la al menos una trasposasa es una trasposasa Tn5. En algunas realizaciones, la al menos una composición de extremos del trasposón incluye el extremo del trasposón Tn5.

En algunas realizaciones, la cadena transferida incluye dominios de marcador que contienen uno o más de un dominio de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección y un dominio de marcador de dirección.

20 En algunas realizaciones, el método proporcionado en el presente documento incluye además (d) incubar la mezcla de la etapa (c) directamente con al menos una enzima modificadora de ácido nucleico en condiciones en donde un marcador de 3' se une a los fragmentos de ácido nucleico objetivo marcados en 5' para generar una pluralidad de fragmentos de ácido nucleico objetivo con dos marcadores. En algunas realizaciones, las etapas (a), (b), (c) y (d) se realizan en un solo tubo de reacción.

25 En algunas realizaciones, la enzima modificadora de ácido nucleico es una polimerasa y en donde dicho marcador de 3' se forma por extensión del extremo 3' del fragmento de ácido nucleico marcado en 5'. En algunas realizaciones, la enzima modificadora de ácido nucleico es una ligasa y en donde el marcador de 3' se forma por ligadura de un oligonucleótido al extremo 3' del fragmento de ácido nucleico marcado en 5'.

30 En algunas realizaciones, el método proporcionado en el presente documento incluye además (e) amplificar uno o más fragmentos de ácido nucleico objetivo con dos marcadores para generar una biblioteca de fragmentos de ácido nucleico marcados con secuencia adicional en el extremo 5' y/o el extremo 3' de los fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, las etapas (a), (b), (c), (d) y (e) se realizan en un solo tubo de reacción.

35 En algunas realizaciones, la amplificación incluye el uso de una o más de una reacción en cadena de la polimerasa (PCR), una reacción de amplificación por desplazamiento de cadena, una reacción de amplificación por círculo rodante, una reacción en cadena de ligasa, una reacción de amplificación mediada por transcripción, o una reacción de amplificación mediada por bucle. En algunas realizaciones, la amplificación incluye una PCR que usa un único cebador que es complementario al marcador de 3' de los fragmentos de ADN objetivo con dos marcadores. En algunas realizaciones, la amplificación incluye una PCR que usa un primer y un segundo cebador, en donde al menos una porción del extremo 3' del primer cebador es complementaria de al menos una porción del marcador 3' de los fragmentos de ácido nucleico objetivo con dos marcadores, y en donde al menos una porción del extremo 3' del segundo cebador presenta la secuencia de al menos una porción del marcador de 5' de los fragmentos de ácido nucleico objetivo con dos marcadores. En algunas realizaciones, una porción del extremo 5' del primer cebador no es complementaria del marcador de 3' de los fragmentos de ácido nucleico objetivo con dos marcadores, y una porción del extremo 5' del segundo cebador no presenta la secuencia de al menos una porción del marcador 5' de los fragmentos de ácido nucleico objetivo con dos marcadores. En algunas realizaciones, el primer cebador incluye una primera secuencia universal, y/o en donde el segundo cebador incluye una segunda secuencia universal.

45 En algunas realizaciones, el método proporcionado en el presente documento incluye además secuenciar los fragmentos de ácido nucleico marcados. En algunas realizaciones, la secuenciación de los fragmentos de ácido nucleico marcados incluye el uso de una o más de secuenciación por síntesis, PCR puente, secuenciación por terminación de cadena, secuenciación por hibridación, secuenciación por nanoporos y secuenciación por ligadura. En algunas realizaciones, la secuenciación de los fragmentos de ácido nucleico marcados incluye el uso de secuenciación de próxima generación.

En algunas realizaciones, el método proporcionado en el presente documento incluye además analizar la variación del número de copias. En algunas realizaciones, el método proporcionado en el presente documento incluye además analizar la variación de un solo nucleótido.

55 En otro aspecto, la presente descripción proporciona un kit para preparar una biblioteca de fragmentos de ácido nucleico marcados, que no es parte de la invención. El kit incluye (a) un reactivo de lisis que tiene una o más proteasas,

y (b) una composición de reacción de transposición que tiene al menos una transposasa y al menos una composición de extremos de transposón que contiene una cadena transferida.

5 En algunas realizaciones, la una o más proteasas se seleccionan de un grupo que consiste en serina proteasas, treonina proteasas, cisteína proteasas, aspartato proteasas, ácido glutámico proteasas y metaloproteasas. En algunas realizaciones, la una o más proteasas son subtilisinas y variantes de las mismas. En algunas realizaciones, el agente de lisis incluye uno o más detergentes. En algunas realizaciones, el uno o más detergentes incluyen Triton.

10 En algunas realizaciones, la al menos una composición de extremos de transposón incluye un dominio de marcador y una porción 3' que comprende la cadena transferida. En algunas realizaciones, el dominio de marcador incluye uno o más de un dominio de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección y un dominio de marcador de dirección. En algunas realizaciones, la composición de la reacción de transposición incluye dos o más composiciones de extremos de transposón, cada una de las dos o más composiciones de extremos de transposón incluye una cadena transferida que difiere en al menos un nucleótido. En algunas realizaciones, la transposasa es una transposasa Tn5. En algunas realizaciones, la composición de extremos de transposón incluye un extremo del transposón Tn5.

15 En algunas realizaciones, el kit proporcionado en el presente documento incluye además una polimerasa. En algunas realizaciones, el kit proporcionado en el presente documento incluye además una ligasa.

20 En algunas realizaciones, el kit proporcionado en el presente documento incluye además un reactivo para una reacción de amplificación. En algunas realizaciones, el reactivo para la reacción de amplificación es un reactivo para PCR. En algunas realizaciones, el reactivo para la reacción de amplificación incluye al menos un cebador. En algunas realizaciones, el al menos un cebador incluye una porción 3' que presenta la secuencia de al menos una porción de la cadena transferida. En algunas realizaciones, el al menos un cebador incluye una porción 5' que contiene una secuencia universal.

25 En algunas realizaciones, el kit proporcionado en el presente documento incluye además un reactivo de selección de tamaño. En algunas realizaciones, el reactivo de selección de tamaño incluye perlas AMPure XP. En algunas realizaciones, el kit proporcionado en el presente documento incluye además un reactivo de normalización de biblioteca.

30 En algunas realizaciones, el kit proporcionado en el presente documento incluye además un aparato que tiene una superficie sólida. En algunas realizaciones, el aparato es un aparato de celda de flujo. En algunas realizaciones, la superficie sólida incluye una superficie grabada adecuada para la inmovilización de una molécula en un patrón ordenado.

Breve descripción de los dibujos

La Figura 1 es un histograma que muestra el porcentaje de lecturas mapeadas únicas en una secuencia usando 0 mg/ml, 0.1 mg/ml, 0.5 mg/ml o 2.5 mg/ml de células enteras o núcleos tratados con proteasas.

35 La Figura 2 muestra histogramas de resultados del análisis de número de copias usando ADN en masa, célula individual tratada con suficiente actividad de proteasa y célula individual tratada con insuficiente actividad de proteasa.

40 La Figura 3A muestra histogramas de resultados del análisis de número de copias en una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 2 mg/ml o proteasa inactivada por calor previamente 2 mg/ml. La Figura 3B muestra un histograma de porcentaje de lecturas mapeadas únicas en una secuenciación de una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml, proteasa bajo temperatura de reacción 2 mg/ml, o proteasa inactivada por calor previamente 2 mg/ml y una muestra de control sin células. La Figura 3C muestra un histograma de las diferencias de recuento de lecturas entre clases adyacentes (Intervalo intercuartílico de diferencia de recuentos de lecturas entre clases adyacentes) en una secuenciación de una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml, proteasa activa 2 mg/ml o proteasa inactivada por calor previamente 2 mg/ml, y una muestra de control sin células.

45 La Figura 4A es un histograma que muestra la actividad relativa de la proteasa a pH 7.0, pH 7.5, pH 8.0, pH 8.5, pH 9.0 o pH 10.0. La Figura 4B muestra un histograma de porcentaje de lecturas mapeadas únicas en un experimento de secuenciación de una sola célula tratada con proteasa a pH 7.0, pH 8.0, pH 9.0 o pH 10.0. La Figura 4C muestra un histograma de las diferencias de recuentos de lecturas entre clases adyacentes (Intervalo intercuartílico de diferencia de recuentos de lecturas entre clases adyacentes) en un experimento de secuenciación de una sola célula tratada con proteasa a pH 7.0, pH 8.0, pH 9.0 o pH 10.0.

50 La Figura 5A es un histograma que muestra la actividad relativa de proteasa precalentada a temperatura ambiente, 50°C, 60°C o 70°C. La Figura 5B muestra un histograma de porcentaje de lecturas mapeadas únicas en un experimento de secuenciación de una sola célula, tres células o 15 pg de ADN genómico, tratados con proteasa precalentada a temperatura ambiente, 50°C, 60°C o 70°C. La Figura 5C muestra un histograma de diferencias de recuentos de lecturas entre clases adyacentes (Intervalo intercuartílico de diferencia de recuentos de lectura entre

clases adyacentes) en un experimento de secuenciación de una sola célula, tres células o 15 pg de ADN genómico, tratado con proteasa precalentada a temperatura ambiente, 50°C, 60°C o 70°C.

5 La Figura 6A muestra el tamaño del inserto de una biblioteca generada con tratamiento de 1 µl de Tn5 o 2 µl de Tn5. La Figura 6B muestra el tamaño del inserto de una biblioteca generada con tratamiento de 1 µl de Tn5 o 2 µl de Tn5. La Figura 6C muestra la diversidad de las bibliotecas generadas con el tratamiento de 1 µl de Tn5 o 2 µl de Tn5.

La Figura 7 muestra histogramas de recuentos y resultados de análisis de número de copias en un experimento de secuenciación de una sola célula de acuerdo con el método proporcionado en el presente documento usando PCR con 16 ciclos, 18 ciclos o 20 ciclos.

10 La Figura 8A muestra la distribución de lecturas de tres experimentos de secuenciación de células individuales. La Figura 8B muestra la distribución de lecturas de secuenciación de una sola célula, secuenciación de tres células o secuenciación de cinco células. La Figura 8C muestra histogramas de diversidad promedio de bibliotecas y cobertura genómica estimada usando una sola célula, tres células o cinco células. La Figura 8D muestra la tasa de éxito general del protocolo.

15 La Figura 9A muestra el análisis del número de copias usando REPLig Single Cell (MDA) con la preparación de la biblioteca Nextera XT. La Figura 9B muestra el análisis del número de copias usando SurePlex con la preparación de la biblioteca Nextera XT. La Figura 9C muestra el análisis del número de copias usando Nextera de una sola célula proporcionado en el presente documento.

20 La Figura 10A muestra datos de análisis del número de copias del cromosoma 18 usando tres replicas de una sola célula GM50121. La Figura 10B muestra los datos del número de recuentos utilizando tres replicas de una sola célula GM20916. La Figura 10C muestra datos de análisis de número de copias de los cromosomas 15, X y 10 usando tres replicas de una sola célula GM20916. La Figura 10D muestra datos de análisis de número de copias de los cromosomas 1 y 11 usando tres réplicas de una sola célula GM10239.

Descripción detallada

25 La presente descripción se refiere generalmente a métodos para preparar una biblioteca de fragmentos de ácido nucleico, y más específicamente a métodos para preparar una biblioteca de fragmentos de ácido nucleico en una sola mezcla de reacción, p. ej., un solo tubo, que usan proteasas para una variedad de aplicaciones que incluyen, p. ej., secuenciación de próxima generación.

Definiciones

30 Como se usa en el presente documento, los términos "incluye", "que incluye", "incluye", "que incluye", "contiene", "que contiene", "tiene", "que tiene" y cualquier variación de los mismos, pretenden cubrir una inclusión no exclusiva, de modo que un procedimiento, método, producto por procedimiento o composición de materia que incluye, incluye o contiene un elemento o lista de elementos no incluye solo esos elementos sino que puede incluir otros elementos que no están expresamente listados o son inherentes a dicho procedimiento, método, producto por procedimiento o composición de la materia.

35 Como se usa en el presente documento, los términos "un" y "una" y "el" y "la" y referentes similares en el contexto de la descripción de la invención (especialmente en el contexto de las siguientes reivindicaciones) deben interpretarse para cubrir tanto el singular como el plural, a menos que se indique lo contrario en el presente documento o que el contexto lo contradiga claramente.

40 Como se usa en el presente documento, el término "alrededor" o "aproximadamente" significa dentro de 5% de un valor o intervalo dado.

45 Como se usa en el presente documento, la expresión "una población mínima de células" significa una población de células que contiene una cantidad de copias de ADN que está por debajo de las capacidades de secuenciación de ácido nucleico en ausencia de una etapa de separación, tal como la extracción de ADN antes de tagmentación. Las etapas de separación de ejemplo incluyen la extracción del contenido de ADN de un lisado celular y/o amplificación de ADN. Una población mínima de células puede incluir una, dos, tres, cuatro o cinco células. Una población mínima de células puede ser una sola célula. "Capacidades de secuenciación de ácido nucleico", como se usa en el presente documento, significa capacidad de secuenciación que puede producir datos de variación de número de copias limpios de un genoma.

50 Como se usa en el presente documento, la expresión "ácido nucleico" significa polímeros monocatenarios y bicatenarios de monómeros nucleótidos, que incluyen 2'-desoxirribonucleótidos (ADN) y ribonucleótidos (ARN) unidos por uniones de enlaces fosfodiéster internucleótidos, o análogos internucleótidos, y los contraiones asociados, p. ej., H⁺, NH₄⁺, trietilamonio, tetraalquilamonio, Mg²⁺, Na⁺ y similares. Un ácido nucleico incluye polinucleótido y oligonucleótido. Un ácido nucleico puede estar compuesto enteramente de desoxirribonucleótidos, enteramente de ribonucleótidos, o mezclas quiméricas de los mismos. Las unidades de monómeros nucleótidos pueden incluir cualquiera de los nucleótidos descritos en el presente documento, incluyendo, pero no limitados a nucleótidos de

origen natural y análogos de nucleótidos. El ácido nucleico típicamente varía en tamaño desde unas pocas unidades monoméricas, p. ej., 5-40, hasta varios miles de unidades de nucleótidos monoméricos. Los ácidos nucleicos incluyen, pero no se limitan a ADN genómico, ADN, ADNc, ARNhn, ARNm, ARNr, ARNt, ácido nucleico fragmentado, ácido nucleico obtenido de orgánulos subcelulares tales como mitocondrias o cloroplastos, y ácido nucleico obtenido de microorganismos o ADN o virus de ARN que pueden estar presentes sobre o en una muestra biológica.

Como se usa en el presente documento, la expresión "ácido nucleico objetivo" pretende significar un ácido nucleico que es objeto de un análisis o acción. El análisis o acción incluye someter el ácido nucleico a copia, amplificación, secuenciación y/u otro procedimiento para la investigación de ácido nucleico. Un ácido nucleico objetivo puede incluir secuencias de nucleótidos adicionales a la secuencia objetivo a analizar. Por ejemplo, un ácido nucleico objetivo puede incluir uno o más adaptadores, incluyendo un adaptador que funciona como un sitio de unión del cebador, que flanquea una secuencia de ácido nucleico objetivo que se analizará. Un ácido nucleico objetivo hibridado con un oligonucleótido de captura o cebador de captura puede contener nucleótidos que se extienden más allá del extremo 5' o 3' del oligonucleótido de captura de tal manera que no todo el ácido nucleico objetivo sea susceptible de extensión.

Como se usa en el presente documento, los términos "aislar" y "purificar" como se usan en el presente documento, se refieren a la reducción en la cantidad de al menos un contaminante (tal como una secuencia de proteína y/o ácido nucleico) de una muestra o de una fuente de la cual se aísla o purifica el material.

Como se usa en el presente documento, el término "selección de tamaño" significa un procedimiento durante el cual se selecciona una subpoblación de fragmentos de ácido nucleico, la mayoría de los cuales tienen una serie de nucleótidos que caen en un intervalo definido, de una población de fragmentos de ácido nucleico, y por lo tanto aumenta el porcentaje de fragmentos de ácido nucleico que tienen una serie de nucleótidos que caen en el intervalo definido.

Como se usa en el presente documento, el término "proteasa" se refiere a una proteína, polipéptido o péptido que presenta la capacidad de hidrolizar polipéptidos o sustratos que tienen una porción de polipéptido. La o las proteasas proporcionadas en los presentes métodos pueden ser una proteasa única que posee una amplia especificidad. Los presentes métodos pueden usar una mezcla de varias proteasas. Las proteasas proporcionadas en el presente documento pueden ser termolábiles y, por lo tanto, pueden inactivarse por calor. En ciertas realizaciones, las proteasas proporcionadas en el presente documento pueden inactivarse a una temperatura superior a aproximadamente 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, 70°C, 75°C, 80°C o superior a aproximadamente 85°C. Las proteasas proporcionadas en el presente documento pueden digerir proteínas de la cromatina y otras proteínas de unión al ADN para liberar ADN genómico desnudo, y también pueden digerir la ADNasa endógena para proteger el ADN de la degradación. Las proteasas proporcionadas en el presente documento incluyen, pero no se limitan a serina proteasas, treonina proteasas, cisteína proteasas, aspartato proteasas, ácido glutámico proteasas y metaloproteasas. Típicamente, las proteasas aspárticas, glutámicas y metaloproteasas activan una molécula de agua que realiza un ataque nucleófilo en el enlace peptídico para hidrolizarlo. Las proteasas de serina, treonina y cisteína usan típicamente un residuo nucleófilo para realizar un ataque nucleófilo para unir covalentemente la proteasa a la proteína del sustrato, liberando la primera mitad del producto. Este compuesto intermedio covalente de acil-enzima después es hidrolizado por agua activada para completar la catálisis liberando la segunda mitad del producto y regenerando la enzima libre. La proteasa de ejemplo usada en el presente documento incluye una serina proteasa aislada de una cepa de *Bacillus recombinante*. Las proteasas de ejemplo usadas en el presente documento incluyen subtilisina y variantes de la misma, que incluyen subtilisina Carlsberg, alcalasa y subtilisina S41. Los expertos en la técnica conocen subtilisinas y variantes de las mismas e incluyen, por ejemplo, alcalasa, alcalasa 0.6L, alcalasa 2.5L, enzima ALK, bacilopeptidasa A, bacilopeptidasa B, proteinasa alcalina bioprasa de *Bacillus subtilis*, bioprasa AL 15, bioprasa APL 30, colistinasa, subtilisina J, subtilisina S41, subtilisina Sendai, subtilisina GX, subtilisina E, subtilisina BL, genenasa I, esperasa, maxatasa, termoasa PC 10, proteasa XXVII, termoasa, superasa, subtilisina Carlsberg subtilisina DY, subtilopeptidasa, SP 266, savinasa 8.0L, savinasa 4.0T, kazusasa, proteasa VIII, opticlean, protina A 3L, savinasa, savinasa 16.0L, savinasa 32.0 L EX, orientasa 10B, proteasa S, serina endopeptidasa. En realizaciones particulares de los métodos y composiciones presentados en el presente documento, se puede usar una proteasa termolábil tal como la subtilisina y las variantes termolábiles de subtilisina, tal como se representa mediante la descripción de ejemplo de Davail et al., 1994, *J. Biol. Chem.*, 26: 17448-17453.

Como se usa en el presente documento, la expresión "inhibidor de proteasa" se refiere a una sustancia, p. ej., un compuesto, capaz de reducir al menos parcialmente la capacidad de una proteasa para hidrolizar péptidos.

Como se usa en el presente documento, el término "ligasa" se refiere a una enzima modificadora de ácido nucleico que cataliza la formación intra e intermolecular de enlaces fosfodiéster entre los extremos 5'-fosfato y 3'-hidroxilo de cadenas de ácido nucleico. Las ligasas incluyen, p. ej., ligasas independientes de la plantilla, tales como la ADNmc ligasa CIRCLIGASE™, que puede unir los extremos de ARN y ADN monocatenarios, y dependientes de la plantilla, que sellan mellas en el ADN bicatenario. Como se usa en el presente documento, "ligasa dependiente de plantilla" significa una ADN ligasa que cataliza la formación intra e intermolecular de enlaces fosfodiéster entre los extremos 5'-fosfato y 3'-hidroxilo de las cadenas de ADN que son adyacentes entre sí cuando se reasocian con un polinucleótido complementario. El polinucleótido al que ambos de los extremos de ADN se van a reasociar por ligación de forma adyacente se denomina en el presente documento "plantilla de ligadura" y la ligadura se denomina "ligadura dependiente de plantilla". La plantilla de ligadura puede ser una secuencia de ADN complementaria en el ADN

genómico u otro en una muestra biológica, o la plantilla de ligadura puede ser un "oligodesoxirribonucleótido puente" o una "férula de ligadura oligodesoxirribonucleótido" (o "férula de ligadura") que se sintetiza y/o proporciona específicamente para usar en un ensayo o método particular. Ejemplos de ADN ligasas dependientes de plantilla incluyen ADN ligasas tipo NAD tales como ADN ligasa de *E. coli*, ADN ligasa Tth, ADN ligasa Tfl y ADN ligasa AMPLIGASE® (EPICENTER Biotechnologies, Madison, WI, EE.UU.), que catalizan la ligadura intramolecular de moléculas de ADNmc solo en presencia de una plantilla de ligadura, y ADN ligasas de tipo ATP, tal como la ADN ligasa T4 o la ADN ligasa FASTLINK™ (EPICENTER Biotechnologies).

Como se usa en el presente documento, el término "tagmentación" se refiere a la modificación del ADN por un complejo de transposoma que comprende la enzima transposasa complejada con adaptadores que comprenden secuencia de extremo del transposón. La tagmentación da como resultado la fragmentación simultánea del ADN y la ligadura de los adaptadores a los extremos 5' de ambas cadenas de fragmentos dúplex. Se pueden añadir secuencias adicionales a los extremos de los fragmentos adaptados, por ejemplo, mediante PCR, ligadura o cualquier otra metodología adecuada conocida por los expertos en la técnica. Como se usa en el presente documento, la expresión "complejo de transposoma" se refiere a una enzima transposasa unida de forma no covalente a un ácido nucleico bicatenario. Por ejemplo, el complejo puede ser una enzima transposasa preincubada con ADN bicatenario de transposón en condiciones que soportan la formación de complejo no covalente. El ADN bicatenario de transposón puede incluir, sin limitación, ADN de Tn5, una porción de ADN de Tn5, una composición de extremo de transposón, una mezcla de composiciones de extremos de transposón u otros ADN bicatenarios capaces de interactuar con una transposasa tal como la transposasa Tn5 hiperactiva.

Como se usa en el presente documento, la expresión "reacción de transposición" se refiere a una reacción en donde uno o más transposones se insertan en ácidos nucleicos objetivo, p. ej., en sitios aleatorios o sitios casi aleatorios. Los componentes esenciales en una reacción de transposición son una transposasa y oligonucleótidos de ADN que presenten las secuencias de nucleótidos de un transposón, incluyendo la secuencia del transposón transferida y su complemento (la secuencia de extremo del transposón no transferida), así como otros componentes necesarios para formar una transposición funcional o complejo de transposoma. Los oligonucleótidos de ADN pueden incluir además secuencias adicionales (p. ej., secuencias de adaptador o cebador) según sea necesario o deseado. En algunas realizaciones, el método proporcionado en el presente documento se ilustra empleando un complejo de transposición formado por una transposasa Tn5 hiperactiva y un extremo de transposón de tipo Tn5 (Goryshin y Reznikoff, 1998, *J. Biol. Chem.*, 273: 7367) o por una transposasa MuA y un extremo de transposón Mu que comprende secuencias de extremo R1 y R2 (Mizuuchi, 1983, *Cell*, 35: 785; Savilahti et al., 1995, *EMBO J.*, 14: 4893). Sin embargo, cualquier sistema de transposición que sea capaz de insertar un extremo de transposón de manera aleatoria o casi aleatoria con suficiente eficiencia para marcar 5' y fragmentar un ADN objetivo para su propósito previsto puede usarse en la presente invención. Los ejemplos de sistemas de transposición conocidos en la técnica que se pueden usar para los presentes métodos incluyen, pero no se limitan a *Staphylococcus aureus* Tn552 (Colegio et al., 2001, *J. Bacteriol.*, 183: 2384-8; Kirby et al., 2002, *Mol Microbiol*, 43: 173-86), Tyl (Devine y Boeke, 1994, *Nucleic Acids Res.*, 22: 3765-72 y Solicitud de Patente Internacional N° WO 95/23875), Transposón Tn7 (Craig, 1996, *Science*, 271: 1512; Craig, 1996, Revisión en: *Curr Top Microbiol Immunol*, 204: 27-48), TnIO e ISIO (Kleckner et al., 1996, *Curr Top Microbiol Immunol*, 204: 49-82), Transposasa Mariner (Lampe et al., 1996, *EMBO J.*, 15: 5470-9), Tci (Plasterk, 1996, *Curr Top Microbiol Immunol*, 204: 125-43), Elemento P (Gloor, 2004, *Methods Mol Biol*, 260: 97-114), TnJ (Ichikawa y Ohtsubo, 1990, *J Biol Chem*, 265: 18829-32), secuencias de inserción bacteriana (Ohtsubo y Sekine, 1996, *Curr. Top. Microbiol. Immunol.* 204: 1-26), retrovirus (Brown et al., 1989, *Proc Natl Acad Sci USA*, 86: 2525-9) y retrotransposón de levadura (Boeke y Corces, 1989, *Annu Rev Microbiol.* 43: 403-34). El método para insertar un extremo de transposón en una secuencia objetivo puede llevarse a cabo in vitro utilizando cualquier sistema de transposón adecuado para el cual esté disponible un sistema de transposición in vitro adecuado o que pueda desarrollarse basándose en el conocimiento en la técnica. En general, un sistema de transposición in vitro adecuado para usar en los métodos proporcionados en el presente documento requiere, como mínimo, una enzima transposasa de pureza suficiente, concentración suficiente y suficiente actividad de transposición in vitro y un extremo de transposón con el que la transposasa forma un complejo funcional con la respectiva transposasa que es capaz de catalizar la reacción de transposición. Las secuencias de extremos de transposón de transposasa adecuadas que se pueden usar en la invención incluyen, pero no se limitan a secuencias de extremos de transposón de tipo natural, derivadas o mutantes que forman un complejo con una transposasa elegida entre una forma de tipo natural, derivada o mutante de la transposasa.

Como se usa en el presente documento, el término "transposasa" se refiere a una enzima que es capaz de formar un complejo funcional con una composición que contiene extremos de transposón (p. ej., transposones, extremos de transposones, composiciones de extremos de transposones) y catalizar la inserción o transposición de la composición que contiene extremos de transposón en el ácido nucleico bicatenario objetivo con el que se incubaba, por ejemplo, en una reacción de transposición in vitro. Una transposasa como se presenta en el presente documento también puede incluir integrasas de retrotransposones y retrovirus. Las transposasas, los transposomas y los complejos de transposomas son generalmente conocidos por los expertos en la técnica, como se ilustra por la descripción del documento US 2010/0120098. Aunque muchas realizaciones descritas en el presente documento se refieren a la transposasa Tn5 y/o la transposasa Tn5 hiperactiva, se apreciará que cualquier sistema de transposición que sea capaz de insertar un extremo de transposón con suficiente eficacia al marcador de 5' y fragmentar un ácido nucleico objetivo para su propósito pretendido puede usarse en la presente invención. En realizaciones particulares, un sistema de transposición es capaz de insertar el extremo del transposón de manera aleatoria o casi aleatoria en el marcador

de 5' y fragmentar el ácido nucleico objetivo.

Como se usa en el presente documento, la expresión "extremo de transposón" significa un ADN bicatenario que presenta solo las secuencias de nucleótidos (las "secuencias de extremos de transposón") que son necesarias para formar el complejo con la enzima transposasa o integrasa que es funcional en una reacción de transposición in vitro.

5 Un extremo de transposón forma un "complejo" o un "complejo sináptico" o un "complejo de transposoma" o una "composición de transposoma con una transposasa o integrasa que reconoce y se une al extremo del transposón, y cuyo complejo es capaz de insertar o transponer el extremo de transposón en el ADN objetivo con el que se incubaba en una reacción de transposición in vitro. Un extremo del transposón presenta dos secuencias complementarias que consisten en una "cadena transferida" y una "cadena no transferida". Por ejemplo, un extremo del transposón que
10 forma un complejo con un transposasa hiperactiva Tn5 (p. ej., Transposasa EZ-Tn5, EPICENTER Biotechnologies, Madison, WI, EE.UU.) que es activa en una reacción de transposición in vitro comprende una cadena transferida que presenta una "secuencia de extremo de transposón transferida" como sigue: 5' AGATGTGTATAAGAGACAG 3' (SEQ ID NO: 1) y una cadena no transferida que presenta una "secuencia de extremo de transposón no transferida" como sigue: 5' CTGTCT CTTATACATCT 3' (SEQ ID NO: 2). El extremo 3' de una cadena transferida se une o transfiere
15 al ácido nucleico objetivo en una reacción de transposición in vitro. La cadena no transferida, que presenta una secuencia de extremo de transposón que es complementaria a la secuencia de extremo del transposón transferido, no se une o transfiere al ácido nucleico objetivo en una reacción de transposición in vitro.

Como se usa en el presente documento, la expresión "composición de extremos de transposón" se refiere a una composición que comprende un extremo del transposón (el segmento mínimo de ADN bicatenario que es capaz de actuar con una transposasa para experimentar una reacción de transposición), opcionalmente más secuencia o secuencias adicionales. 5' de la secuencia de extremo de transposón transferido y/o 3' de la secuencia de extremo de transposón no transferido. Por ejemplo, un extremo de transposón unido a un marcador es una "composición de extremos de transposón".

Como se usa en el presente documento, la expresión "cadena transferida" se refiere a la porción transferida tanto de "extremos del transposón" como de "composiciones de extremos de transposón" (independientemente de si el extremo del transposón está unido a un marcador u otro resto). De manera similar, la expresión "cadena no transferida" se refiere a la porción no transferida tanto de "extremos de transposón" como de "composiciones de extremos de transposón".

Como se usa en el presente documento, el término "marcador" se refiere a un componente de ácido nucleico no objetivo, generalmente ADN, que proporciona un medio para dirigirse a un fragmento de ácido nucleico al que se une. Por ejemplo, en algunas realizaciones, un marcador comprende una secuencia de nucleótidos que permite la identificación, el reconocimiento y/o la manipulación molecular o bioquímica del ADN al que está unido el marcador (p. ej., proporcionando un sitio para la reasociación de un oligonucleótido, tal como un cebador para la extensión por una ADN polimerasa, o un oligonucleótido para capturar o para una reacción de ligadura). El procedimiento de unir el marcador a la molécula de ácido nucleico a veces se denomina en el presente documento "marcaje" y el ácido nucleico que es marcado o que contiene un marcador se denomina "marcado" (p. ej., "ADN marcado").

Como se usa en el presente documento, la expresión "dominio de marcador" se refiere a una porción o dominio de un marcador que presenta una secuencia para un propósito o aplicación deseada. Un dominio de marcador es el "dominio del extremo del transposón", cuyo dominio de marcador presenta la secuencia del extremo del transposón transferida. En algunas realizaciones, la cadena transferida también presenta una o más secuencias de nucleótidos 5' distintas de la secuencia de extremo de transposón transferida, el marcador también tiene uno o más "dominios de marcador" en la porción 5', cada uno de cuyos dominios de marcador se proporciona para cualquier propósito deseado. Por ejemplo, algunas realizaciones contienen una composición de extremos de transposón que incluye un dominio de marcador seleccionado entre uno o más de un dominio de marcador de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección, un dominio de marcador de dirección y un dominio promotor de transcripción.

Como se usa en el presente documento, la expresión "dominio del sitio de restricción" se refiere a un dominio de marcador que presenta una secuencia con el fin de facilitar la escisión usando una endonucleasa de restricción. Por ejemplo, el dominio del sitio de restricción puede usarse para generar fragmentos de ADNmc lineales con dos marcadores. El dominio del sitio de restricción también se puede usar para generar un extremo 5' bicatenario compatible en el dominio del marcador de modo que este extremo se pueda ligar a otra molécula de ADN usando una ADN ligasa dependiente de plantilla.

Como se usa en el presente documento, la expresión "dominio de marcador de captura" se refiere a un dominio de marcador que presenta una secuencia para el propósito de facilitar la captura del fragmento de ácido nucleico al que se une el dominio de marcador (p. ej., para proporcionar un sitio de reasociación o un marcador de afinidad para una captura de los fragmentos lineales de ADNmc lineales con dos marcadores en la superficie de una perla u otra, p. ej., en donde el sitio de reasociación de la secuencia del dominio de marcador permite la captura por reasociación con una secuencia específica que está en una superficie, tal como una sonda en una perla o en un microchip o micromatrices o en una perla de secuenciación). En algunas realizaciones, el dominio de marcador de captura
60 comprende una porción 5' de la cadena transferida que se une a un grupo químico o resto que incluye una molécula

de unión por afinidad (p. ej., biotina, estreptavidina, un antígeno o un anticuerpo que se une al antígeno, que permite la captura de los fragmentos lineales de ADNmc con dos marcadores en una superficie a la que se une una segunda molécula de unión por afinidad que forma una pareja de unión específica con la primera molécula de unión por afinidad).

5 Como se usa en el presente documento, la expresión "dominio de marcador de secuenciación" se refiere a un dominio de marcador que presenta una secuencia para el propósito de facilitar la secuenciación del fragmento de ácido nucleico al que se une al marcador (p. ej., para proporcionar un sitio de cebado para la secuenciación por síntesis, o para proporcionar sitios de reasociación para la secuenciación por ligadura, o para proporcionar sitios de reasociación para la secuenciación por hibridación).

10 Como se usa en el presente documento, la expresión "dominio de marcador de amplificación" se refiere a un dominio de marcador que presenta una secuencia para el propósito de facilitar la amplificación de un ácido nucleico al que se adjunta dicho marcador. Por ejemplo, en algunas realizaciones, el dominio de marcador de amplificación proporciona un sitio de cebado para una reacción de amplificación de ácido nucleico usando una ADN polimerasa (p. ej., una reacción de amplificación por PCR o una reacción de amplificación por desplazamiento de cadena, o una reacción de amplificación por círculo rodante), o una plantilla de ligadura para el ligado de sondas usando una ligasa dependiente de plantilla en una reacción de amplificación de ácido nucleico (p. ej., una reacción en cadena de ligadura).

15 Como se usa en el presente documento, la expresión "dominio de marcador de detección" se refiere a un dominio de marcador que presenta una secuencia o un resto químico o bioquímico detectable para el propósito de facilitar la detección de los fragmentos de ácido nucleico marcados (p. ej., un colorante visible, fluorescente, quimioluminiscente u otro detectable; una enzima que es detectable en presencia de un sustrato, p. ej., una fosfatasa alcalina con NBT más BCIP o una peroxidasa con un sustrato adecuado; una proteína detectable, p. ej., una proteína verde fluorescente; y una molécula de unión por afinidad que está unida a un resto detectable o que puede formar una pareja de unión por afinidad o una pareja de unión específica con otra molécula de unión por afinidad detectable; o cualquiera de las muchas otras moléculas o sistemas detectables conocidos en la técnica).

20 Como se usa en el presente documento, la expresión "dominio de marcador de dirección" significa un dominio de marcador que presenta una secuencia que permite la identificación de una muestra específica (p. ej., en donde la cadena transferida tiene un dominio de marcador de dirección diferente que presenta una secuencia diferente para cada muestra).

25 Como se usa en el presente documento, los términos "amplificar" o "amplificado" "que amplifica" como se usan en referencia a un ácido nucleico o reacciones de ácidos nucleicos, se refieren a métodos in vitro para hacer copias de un ácido nucleico particular, tal como un ácido nucleico objetivo o un ácido nucleico marcado. Se conocen numerosos métodos para amplificar ácidos nucleicos en la técnica, y las reacciones de amplificación incluyen, pero no se limitan a reacciones en cadena de la polimerasa, reacciones en cadena de la ligasa, reacciones de amplificación por desplazamiento de cadena, reacciones de amplificación por círculo rodante. El ácido nucleico que se amplifica puede ser ADN. Los productos resultantes de la amplificación de una molécula o moléculas de ácido nucleico ("productos de amplificación"), ya sea que el ácido nucleico inicial sea ADN, ARN o ambos, pueden ser ADN o ARN, o una mezcla de nucleósidos o nucleótidos tanto de ADN como de ARN, o pueden incluir nucleósidos o nucleótidos de ADN o ARN modificados. Una "copia" no significa necesariamente una secuencia perfecta complementaria o idéntica con la secuencia objetivo. Por ejemplo, las copias pueden incluir análogos de nucleótidos tales como la desoxinosina o la desoxiuridina, alteraciones de secuencia deliberadas (tales como las alteraciones de secuencia introducidas a través de un cebador que contiene una secuencia que es hibridable, pero no complementaria, con la secuencia objetivo), y/o errores de secuencia que ocurren durante la amplificación.

30 Como se usa en el presente documento, la expresión "biblioteca de fragmentos de ácido nucleico marcados" se refiere a una colección o población de fragmentos de ácido nucleico marcados (p. ej., fragmentos de ácido nucleico con dos marcadores) generados a partir de un recurso, p. ej., genoma completo, en donde la combinación de los fragmentos de ácido nucleico marcados en la colección o población presenta secuencias que son cualitativa y/o cuantitativamente representativas de la secuencia del recurso a partir del cual se generaron los fragmentos de ácido nucleico marcados, p. ej., todo el genoma. Es posible que una biblioteca de fragmentos de ácido nucleico marcados no contenga un fragmento nucleico marcado que represente cada secuencia que presenta el recurso.

35 Como se usa en el presente documento, la expresión "enzima modificadora de ácido nucleico" se refiere a cualquier enzima que actúa sobre el ácido nucleico, p. ej., ADN, para efectuar una modificación, p. ej., escisión, ligadura, polimerización, fosforilación, etc. Las enzimas modificadoras de ácido nucleico incluyen, p. ej., polimerasas, nucleasas, transferasas, ligasas, fosforilasas, fosfatasa, metilasas, transosasas, etc. Las "enzimas modificadoras de ADN" incluyen cualquier enzima que actúe sobre el ADN, incluyendo enzimas que también actúan sobre otros sustratos, tal como el ARN.

40 Como se usa en el presente documento, la expresión "ADN polimerasa" se refiere a una enzima modificadora que cataliza la polimerización de desoxirribonucleótidos en una cadena de ADN. Las ADN polimerasas incluyen "ADN polimerasas dependientes de plantilla", que requieren un ácido nucleico de plantilla para determinar el orden en que se añaden los desoxirribonucleótidos en el polímero, o pueden ser "independientes de plantilla" de manera que

catalicen la polimerización sin referencia a una secuencia de plantilla. Además de sintetizar polímeros de ADN, las ADN polimerasas pueden comprender otras características o actividades. Por ejemplo, una ADN polimerasa puede caracterizarse por tener o carecer de actividad de exonucleasa de 5' a 3' (también denominada una actividad de 5' exonucleasa o 5' nucleasa), actividad de exonucleasa 3' a 5' y actividad de desplazamiento de cadena.

5 Como se usa en el presente documento, el término "cebador" es un oligonucleótido ("oligo"), generalmente con un grupo 3'-OH libre que se puede extender mediante una ácido nucleico polimerasa. Para una polimerasa dependiente de plantilla, generalmente al menos la porción 3' del cebador oligo es complementaria a una porción de un ácido nucleico plantilla, al cual el oligo "se une" (o "forma complejo", "se reasocia" o "se hibrida"), mediante enlaces de hidrógeno y otras fuerzas moleculares, a la plantilla para dar un complejo cebador/plantilla para el inicio de la síntesis por una ADN polimerasa, y que se extiende mediante la adición de bases unidas covalentemente unidas a su extremo 3' que son complementarias de la plantilla en el proceso de síntesis de ADN. El resultado es un producto de extensión del cebador.

10 Como se usa en el presente documento, la expresión "secuencia universal" se refiere a una región de secuencia de nucleótidos que es común o compartida por dos o más moléculas de ácido nucleico. Opcionalmente, las dos o más moléculas de ácido nucleico también tienen regiones de diferentes secuencias. Así, por ejemplo, los marcadores de 5' pueden comprender secuencias de ácido nucleico idénticas o universales y los marcadores de 3' pueden comprender secuencias idénticas o universales. Una secuencia universal que puede estar presente en diferentes miembros de una pluralidad de moléculas de ácido nucleico puede permitir la replicación o amplificación de múltiples secuencias diferentes usando un único cebador universal que es complementario de la secuencia universal.

20 Como se usa en el presente documento, las expresiones "superficie sólida", "soporte sólido" y otros equivalentes gramaticales en el presente documento se refieren a cualquier material que sea apropiado o pueda modificarse para que sea apropiado para la unión de un polinucleótido. Los posibles sustratos incluyen, pero no se limitan a vidrio y vidrio modificado o funcionalizado, plásticos (incluyendo acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™, etc.), polisacáridos, nylon o nitrocelulosa, materiales cerámicos, resinas, sílice o materiales basados en sílice, incluyendo silicio y silicio modificado, carbono, metales, vidrios inorgánicos, plásticos, haces de fibras ópticas y una variedad de otros polímeros. En algunas realizaciones, los soportes sólidos y las superficies sólidas se encuentran dentro de un aparato de celda de flujo. En algunas realizaciones, el soporte sólido comprende una superficie grabada adecuada para la inmovilización de complejos de transposomas en un patrón ordenado. Una "superficie grabada" se refiere a una disposición de diferentes regiones en o sobre una capa expuesta de un soporte sólido. En algunas realizaciones, el soporte sólido comprende una matriz de pocillos o depresiones en una superficie. La composición y la geometría del soporte sólido pueden variar con su uso. En algunas realizaciones, el soporte sólido es una estructura plana tal como un portaobjetos, chip, microchip y/o matriz. Como tal, la superficie de un sustrato puede tener la forma de una capa plana. En algunas realizaciones, el soporte sólido comprende una o más superficies de una celda de flujo. La expresión "celda de flujo", como se usa en el presente documento, se refiere a una cámara que comprende una superficie sólida a través de la cual pueden fluir uno o más reactivos fluidos. Se describen ejemplos de células de flujo y sistemas fluidicos relacionados y plataformas de detección que pueden usarse fácilmente en los métodos de la presente descripción, por ejemplo, en Bentley et al., *Nature* 456: 53-59 (2008), WO 04/018497; US 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; US 7,405,281 y US 2008/0108082. En algunas realizaciones, el soporte sólido o su superficie no es plana, tal como la superficie interna o externa de un tubo o recipiente. En algunas realizaciones, el soporte sólido comprende microesferas o perlas. Las "microesferas", "perlas", "partículas" o los equivalentes gramaticales en el presente documento pretenden significar partículas discretas pequeñas hechas de diversos materiales que incluyen, pero no se limitan a, plásticos, materiales cerámicos, vidrio y poliestireno. En ciertas realizaciones, las microesferas son microesferas magnéticas o perlas. Alternativa o adicionalmente, las perlas pueden ser porosas. Los tamaños de las perlas varían desde nanómetros, p. ej. 100 nm, a milímetros, p. ej. 1 mm.

Métodos para preparar una biblioteca de fragmentos de ácido nucleico marcados

La presente descripción se refiere generalmente a métodos para preparar una biblioteca de fragmentos de ácido nucleico, y más específicamente a métodos para preparar una biblioteca de fragmentos de ácido nucleico en una única mezcla de reacción, p. ej., un solo tubo de reacción u otro recipiente, que utiliza proteasas, para una variedad de aplicaciones que incluyen, p. ej., secuenciación de ADN de próxima generación, análisis de variaciones de número de copias y análisis de variaciones de un solo nucleótido.

55 Hay una variedad de métodos y aplicaciones para los cuales es deseable preparar una biblioteca de fragmentos de ácido nucleico a partir de una población mínima de células, p. ej., una sola célula, para diversas aplicaciones tales como la secuenciación de un genoma. Los métodos actuales para preparar una biblioteca de fragmentos de ácido nucleico requieren una etapa de extracción y/o amplificación de ácido nucleico por separado antes de la fragmentación del ADN. Típicamente, las células se procesan primero para generar un lisado celular del cual se extrae y purifica el contenido de ácido nucleico objetivo. Después, en una etapa separada, el ácido nucleico objetivo purificado se somete a fragmentación, por ejemplo, usando el transposoma Nextera disponible en Illumina, Inc (San Diego, CA). Esta etapa separada de extracción de ácido nucleico y la transferencia de muestras entre tubos o recipientes de reacción normalmente malgasta muestra de ácido nucleico objetivo y, por lo tanto, hace que los fragmentos de ácido nucleico preparados sean menos propensos a representar suficientemente todo ácido nucleico objetivo de la muestra. Esta

representación insuficiente se vuelve particularmente desafiante cuando la cantidad de muestra de células es limitada o difícil de obtener. Se han desarrollado algunos métodos para resolver este problema en el caso de la entrada de una sola o pocas células mediante una etapa de preamplificación. Sin embargo, estos métodos no resuelven eficientemente el problema de representación insuficiente y típicamente introducen ruidos altos. La presente descripción proporciona una solución a este problema mediante el uso de una mezcla de reacción única, p. ej., en un solo tubo, con protocolo de adición para generar una biblioteca de fragmentos de ácido nucleico. El método proporcionado en el presente documento integra varias etapas, incluyendo la generación de lisado celular, tagmentación y similares, en un único tubo de reacción, opcionalmente usando uno o más protocolos de adición. En dicho método de adición de un solo tubo, la cantidad de materiales de ácido nucleico de partida de las células se conserva, y la biblioteca generada a partir de ellos puede representar mejor el ácido nucleico objetivo, p. ej., un genoma.

En un aspecto, la presente descripción proporciona un método para preparar una biblioteca de fragmentos de ácido nucleico marcados que incluye (a) poner en contacto una población de células directamente con un reactivo de lisis para generar un lisado celular, en donde el reactivo de lisis tiene una o más proteasas, y en donde el lisado celular contiene un ácido nucleico objetivo; (b) inactivar la una o más proteasas para formar un lisado celular inactivado, y (c) aplicar directamente al menos una transposasa y al menos una composición de extremos de transposón que contiene una cadena transferida al lisado celular inactivado en condiciones en las que el ácido nucleico objetivo y la composición de extremos de transposón experimentan una reacción de transposición para generar una mezcla, en donde: (i) el ácido nucleico objetivo se fragmenta para generar una pluralidad de fragmentos de ácido nucleico objetivo, (ii) la cadena transferida de la composición de extremos de transposón se une a extremos 5' de cada uno de una pluralidad de fragmentos de ácido nucleico objetivo para generar una pluralidad de fragmentos de ácido nucleico objetivo marcados en 5', y (iii) el ácido nucleico objetivo es un ADN bicatenario, y en donde el ácido nucleico objetivo sigue siendo el ADN bicatenario antes de aplicar una transposasa y una composición de extremos de transposón en la etapa (c).

En algunas realizaciones, la muestra celular se pone en contacto directamente con un reactivo de lisis combinado que contiene una o más proteasas y, por lo tanto, las proteasas proporcionadas en el presente documento se pueden poner en contacto directamente con las células intactas. En algunas realizaciones, la muestra celular se pone en contacto con un primer reactivo de lisis que contiene detergentes para generar un primer lisado celular, y luego se añade un segundo reactivo de lisis que contiene una o más proteasas al tubo de reacción que contiene el primer lisado celular. En esta alternativa, las proteasas proporcionadas en el presente documento se ponen en contacto con el lisado celular. El ejemplo 1 proporcionado a continuación ilustra un método para generar un lisado celular que contiene ácido nucleico objetivo. En el Ejemplo 1 y las Tablas 1-3 se ilustran mezclas maestras de lisis de ejemplo que contienen detergente y proteasa QIAGEN (San Diego, CA) (componente N^o 19155).

El material de partida según el método proporcionado en el presente documento puede ser una población mínima de células, con la que los protocolos de secuenciación tradicionales típicamente solo pueden producir datos de secuenciación y datos de variación del número de copias que son ruido debido a la insuficiente representación a lo largo del ácido nucleico objetivo, p. ej., un genoma. En algunas realizaciones, una población mínima de células puede contener una, dos, tres, cuatro o cinco células. En algunas realizaciones, una población mínima de células puede ser menos de 10 células, menos de 15 células, menos de 20 células, menos de 25 células, menos de 30 células, menos de 35 células, menos de 40 células, menos de 45 células, menos de 50 células, menos de 60 células, menos de 70 células, menos de 80 menos, menos de 90 células o menos de 100 células. En una realización, el material de partida usado en el presente método contiene solo una célula única. En algunas realizaciones, el ácido nucleico objetivo es ADN genómico. En algunas realizaciones, el ácido nucleico objetivo contiene ADN cromosómico o un fragmento del mismo. En algunas realizaciones, el ácido nucleico objetivo comprende un genoma o un genoma parcial.

Las proteasas utilizadas en el presente documento pueden digerir proteínas de cromatina, p. ej., histonas y otras proteínas de unión al ADN para liberar ADN genómico desnudo. Además, las proteasas proporcionadas en el presente documento pueden digerir la DNasa endógena para proteger al genoma de la degradación. En algunas realizaciones, el método de la presente memoria utiliza solo una proteasa que tiene una amplia especificidad y, por lo tanto, las proteasas pueden digerir varias proteínas y polipéptidos diferentes que incluyen algunas o muchas de las proteínas en una célula. En algunas otras realizaciones, la amplia especificidad se puede lograr usando una mezcla de varias proteasas, y la combinación de varias proteasas puede digerir varias proteínas y polipéptidos diferentes que incluyen algunas o muchas de las proteínas en una célula. Las proteasas de ejemplo incluyen subtilisinas tales como alcalasa, subtilisina carlsberg, subtilisina S41, proteinasa K termolábil y proteasa Qiagen. El ejemplo 4 ilustra que la actividad de proteasa es útil para el acceso uniforme al ADN genómico. Debe apreciarse que se pueden usar diferentes proteasas y/o mezclas de proteasas dependiendo de diversas condiciones, p. ej., tipo de célula y cantidad de muestra.

La cantidad y concentración de proteasas utilizadas en cada reacción proporcionada en el presente documento puede variar dependiendo de la cantidad de ADN cromosómico y/o el número de células utilizadas, así como la actividad de las proteasas. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 0.1 mg/ml a 10 mg/ml. En alguna realización, la concentración de una o más proteasas en el lisado celular es de 0.1 mg/ml a 2.5 mg/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 2 mUA/ml a 500 mUA/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 4.5 mUA/ml a 500 mUA/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 10 mUA/ml a 100 mUA/ml. La presente descripción ilustra el ensayo y optimización de la concentración de proteasa usando una

proteasa, p. ej., proteasa QIAGEN (Componente N° 19155) como se muestra en el Ejemplo 5. Como se muestra en este ejemplo, cuando una sola célula se trata con 0.5 mg/ml (equivalente a 22.5 mUA/ml) o 2 mg/ml (equivalente a 90 mUA/ml) de proteasa a temperatura de reacción normal (p. ej., temperatura ambiente), el resultado del análisis del número de copias limpias se logra de manera similar como se muestra en los dos histogramas superiores de la Figura 3A. Por lo tanto, en algunas realizaciones, la concentración de las proteasas en el lisado celular es de 0.5 mg/ml a 2 mg/ml. Ejemplos de las concentraciones de las proteasas en el lisado celular incluyen 0.5 mg/ml, 0.6 mg/ml, 0.7 mg/ml, 0.8 mg/ml, 0.9 mg/ml, 1.0 mg/ml, 1.1 mg/ml, 1.2 mg/ml, 1.3 mg/ml, 1.4 mg/ml, 1.5 mg/ml, 1.6 mg/ml, 1.7 mg/ml, 1.8 mg/ml, 1.9 mg/ml y 2.0 mg/ml. En algunas realizaciones, la concentración de una o más proteasas en el lisado celular es de 20 mUA/ml a 90 mUA/ml. Las concentraciones de ejemplo de una o más proteasas en el lisado celular incluyen 20 mUA/ml, 30 mUA/ml, 40 mUA/ml, 50 mUA/ml, 60 mUA/ml, 70 mUA/ml, 80 mUA/ml, 90 mUA/ml.

Diversas condiciones, incluido el valor de pH, pueden afectar tanto a la digestión por proteasas como a las sustancias activas de otras enzimas en el tubo de reacción y, por lo tanto, estas condiciones, p. ej., valor de pH, se pueden optimizar. El ejemplo 6 ilustra las condiciones de optimización del pH de la reacción de digestión de proteasa equilibrando la actividad de proteasa y los resultados de secuenciación. Como se muestra, la actividad de la proteasa QIAGEN se analiza bajo diferentes condiciones de pH, y la actividad de la proteasa aumenta a medida que el valor del pH aumenta teniendo la proteasa la actividad más baja a pH 7.0 y la actividad más alta a pH 10.0 entre el intervalo de pH 7.0 a pH 10.0. Luego, el porcentaje de lecturas mapeadas únicas y el ruido en los datos del número de copias también se analizan en varias condiciones de pH. Como se muestra, cuando el pH es 7, 8 o 9, se puede lograr aproximadamente 70% de lecturas mapeadas únicas limpias. Sin embargo, cuando el pH es 10, se puede lograr menos porcentaje de lecturas mapeadas únicas y la variación de datos aumenta significativamente. De manera similar, cuando el pH es de 7, 8 o 9, las diferencias de recuento entre clases adyacentes son relativamente pequeñas (aproximadamente 20%) con pequeñas variaciones; mientras que las diferencias de recuento entre clases adyacentes aumentan significativamente con una gran variación a pH 10.0. Por lo tanto, en algunas realizaciones, la población de células se pone en contacto con el reactivo de lisis de pH 7.0 a pH 10.0. En algunas realizaciones, la población de células se pone en contacto con el reactivo de lisis a pH 7.0 a pH 9.0. Las condiciones de pH de ejemplo incluyen pH 7.0, pH 7.5, pH 8.0, pH 8.5, pH 9.0 y pH 9.5.

Debido a que las etapas de preparación de ácido nucleico y tagmentación se realizan en el mismo tubo de reacción, puede ser beneficioso que las proteasas de acuerdo con el presente método se puedan inactivar efectivamente sin alterar la siguiente etapa de tagmentación que típicamente requiere ADN bicatenario. En algunas realizaciones, las proteasas se pueden inactivar aumentando la temperatura antes de la etapa de tagmentación. La alta temperatura puede desnaturar la conformación de ADN bicatenario. Por lo tanto, en algunas realizaciones, las proteasas proporcionadas en el presente documento pueden inactivarse a temperatura relativamente baja sin desnaturar el ADN bicatenario. El ejemplo 7 ilustra el ensayo de inactivación por calor de una proteasa. Como se muestra, la actividad de la proteasa se prueba a diferentes temperaturas, y la actividad de la proteasa disminuye progresivamente a medida que aumenta la temperatura, y se inactiva por completo a 70°C. Por lo tanto, en algunas realizaciones, una o más proteasas se inactivan aumentando la temperatura a 50°C-80°C. En algunas realizaciones, la una o más proteasas se inactivan aumentando la temperatura a 70°C.

En algunas realizaciones, las proteasas proporcionadas en el presente documento también se pueden inactivar añadiendo inhibidores de proteasas al tubo de reacción. Los inhibidores de la proteasa proporcionados en el presente documento no interfieren con la etapa de tagmentación y amplificación que se llevará a cabo más tarde en el mismo tubo de reacción. Los inhibidores de proteasa de ejemplo incluyen, por ejemplo, AEBSF, bestatina, E-64, pepstatina A, fosforamidón, leupeptina, aprotinina, hidrocloreto de bestatina, leupeptina, sal disódica del fosforamidón, elastatinal, aprotinina, mesilato de nafamostat, antipaína, PMSF (fluoruro de fenilmetanosulfonilo), Pefabloc, fluorofosfato de diisopropilo e inhibidor de subtilisina de Streptomyces.

Como se ha descrito anteriormente, también se pueden añadir uno o más detergentes a las células. En algunas realizaciones, los detergentes se añaden a las células junto con las proteasas. En otras realizaciones, los detergentes se añaden a las células primero seguido de la adición de proteasas al tubo de reacción. La función del detergente usado en el presente documento incluye romper las membranas celulares y liberar materiales intracelulares en una forma soluble. En algunas realizaciones, el detergente usado en el presente documento no interfiere con las actividades enzimáticas posteriores. Por lo tanto, en algunas realizaciones, se usan detergentes no iónicos. Estos detergentes rompen las asociaciones proteína-lípido y lípido-lípido, pero no las interacciones proteína-proteína, y por lo tanto son menos propensos a interferir con otras enzimas aguas abajo. Típicamente, los detergentes no iónicos contienen grupos de cabeza hidrófilos sin carga. Los detergentes no iónicos típicos se basan en polioxietileno o un glucósido. Ejemplos de detergentes no iónicos incluyen Tween® 80, Tween® 20, Triton® X-100, Triton® X-100-R, Triton® X-114, NP-40, Genapol® C-100, Genapol® X-100, Igepal® CA 630, Arlasolve® 200, Brij® 96/97, Triton, series Brij® 98, Brij® 58, Brij® 35, Pluronic® L64, Pluronic® P84, sulfobetainas no detergentes (NDSB 201), anfipoles (PMAL-C8), CHAPS, octil β-D-glucopiranosido, saponina, monododecil éter del nonaetilenglicol (C12E9, polidocenol), dodecilsulfato de sodio, N-laurilsarcosina, desoxicolato de sodio, sales biliares, bromuro de hexadeciltrimetilamonio, SB3-10, SB3-12, amidosulfobetaina-14, octil tioglucósido, maltósidos, HEGA y serie MEGA.

Una vez que las proteasas se inactivan, se puede llevar a cabo una reacción de transposición in vitro en la misma mezcla de reacción. p. ej., en el mismo tubo de reacción, añadiendo una composición de transposoma que contiene un complejo estable formado entre la transposasa y la composición de extremos de transposón o usando transposasa

y una composición de extremos de transposón separados. La reacción de transposición in vitro catalizada por una transposasa da como resultado la ruptura simultánea de un ácido nucleico objetivo en fragmentos y la unión de un marcador en el extremo 5' de cada fragmento. Debe entenderse que cualquier método que describa el uso de una transposasa y una composición de extremos de transposón también podría usar una composición de transposoma hecha a partir de la transposasa y la composición de extremos de transposón, y cualquier método que describa el uso de una composición de transposoma también podría usar la transposasa separada y una composición de extremos de transposón de la que se compone la composición de transposoma.

En algunas realizaciones, el método proporcionado en este documento incluye incubar el lisado celular inactivado que contiene el ácido nucleico objetivo en una reacción de transposición in vitro con al menos una transposasa y una composición de extremos del transposón con la cual la transposasa forma un complejo de transposición, la composición de extremos de transposón incluye (i) una cadena transferida que presenta una secuencia de extremo del transposón transferido y, opcionalmente, una secuencia adicional 5' de la secuencia de extremo del transposón transferido, y (ii) una cadena no transferida que presenta una secuencia que es complementaria a la secuencia de extremo del transposón transferido, en condiciones y durante el tiempo suficiente en el que se producen múltiples inserciones en el ácido nucleico objetivo, cada una de las cuales da como resultado la unión de un primer marcador que contiene la cadena transferida al extremo 5' de un nucleótido en el ácido nucleico objetivo, fragmentando así el ácido nucleico objetivo y generando una población de fragmentos de ADN marcados en 5' reasociados, cada uno de los cuales tiene el primer marcador en el extremo 5' de los fragmentos de ácido nucleico objetivo.

En algunas realizaciones, el método descrito anteriormente se lleva a cabo usando transposasa y composiciones de extremos de transposón separados. En otras realizaciones, el método descrito anteriormente se lleva a cabo usando una composición de transposoma que comprende el complejo formado entre la transposasa y la composición de extremos del transposón.

En algunas realizaciones específicas, el método proporcionado en el presente documento se lleva a cabo usando el Transposoma Nextera disponible en Illumina Inc (San Diego, CA), como se describe generalmente en la descripción del documento US 2010/0120098.

Las transposasas y composiciones de transposomas son generalmente conocidas por los expertos en la técnica, como se ilustra por la descripción del documento US 2010/0120098. En algunas realizaciones, el método proporcionado en el presente documento emplea una composición de transposoma formada por una transposasa Tn5 hiperactiva y un extremo de transposón de tipo Tn5 (Goryshin y Reznikoff, 1998, *J. Biol. Chem.*, 273: 7367). En algunas realizaciones, el método proporcionado en el presente documento emplea una composición de transposoma formada o por una transposasa de MuA y un extremo de transposón de Mu que comprende secuencias de extremos R1 y R2 (Mizuuchi, 1983, *Cell*, 35: 785; Savilahti et al., 1995, *EMBO J.*, 14: 4893). Cualquier sistema de transposición que sea capaz de insertar un extremo de transposón de manera aleatoria o casi aleatoria con suficiente eficiencia para marcar en 5' y fragmentar un ácido nucleico objetivo para su propósito previsto puede usarse en la presente descripción. Los ejemplos de sistemas de composición de transposomas incluyen, pero no se limitan a *Staphylococcus aureus* Tn552 (Colegio et al., 2001, *J Bacteriol.*, 183: 2384-8; Kirby et al., 2002, *Mol Microbiol*, 43: 173-86), Tyl (Devine y Boeke, 1994, *Nucleic Acids Res.*, 22: 3765-72 y Solicitud de Patente Internacional N° WO 95/23875), Transposón Tn7 (Craig, 1996, *Science*. 271: 1512; Craig, 1996, Revisión en: *Curr Top Microbiol Immunol*, 204: 27-48), TnIO e ISIO (Kleckner et al., 1996, *Curr Top Microbiol Immunol*, 204: 49-82), Transposasa Mariner (Lampe et al., 1996, *EMBO J.*, 15: 5470-9), Tci (Plasterk, 1996, *Curr Top Microbiol Immunol*, 204: 125-43), Elemento P (Gloor, 2004, *Methods Mol Biol*, 260: 97-114), TnJ (Ichikawa y Ohtsubo, 1990, *J Biol Chem*. 265: 18829-32), secuencias de inserción bacteriana (Ohtsubo y Sekine, 1996, *Curr. Top. Microbiol Immunol* 204: 1-26), retrovirus (Brown et al., 1989, *Proc Natl Acad Sci USA*, 86: 2525-9) y retrotransposón de levadura (Boeke y Corces, 1989, *Annu Rev Microbiol*. 43: 403-34).

Como ejemplos no limitantes, los extremos de transposón pueden incluir el extremo del transposón extremo externo ("OE") de 19 pb, el extremo del transposón extremo interno ("IE") o el extremo del transposón "extremo mosaico" ("ME") reconocido por una transposasa Tn5 de tipo natural o mutante, o el extremo de transposón R1 y R2 como se expone en la descripción del documento US 2010/0120098. Los extremos del transposón pueden incluir cualquier ácido nucleico o análogo de ácido nucleico adecuado para formar un complejo funcional con la transposasa o la enzima integrasa en una reacción de transposición in vitro. Por ejemplo, el extremo del transposón puede incluir ADN, ARN, bases modificadas, bases no naturales, cadena principal modificada, y puede incluir mellas en una o ambas cadenas.

En algunas realizaciones, en donde la cadena transferida incluye una porción 3' y una porción 5', en donde la porción 3' presenta la secuencia de extremo del transposón transferido, y la porción 5' de la cadena transferida presenta una secuencia que comprende uno o más dominios de marcador para un propósito particular (p. ej., un dominio de marcador de secuenciación o un dominio de marcador de amplificación, y opcionalmente un dominio de marcador de dirección para la secuenciación o amplificación de próxima generación). Los dominios de marcador de ejemplo incluyen un dominio de marcador de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección, un dominio de marcador de dirección y un dominio promotor de la transcripción.

En algunas realizaciones, se usan dos transposomas diferentes en la reacción de transposición in vitro, y cada uno de los dos transposomas contiene la misma transposasa pero una composición de extremos de transposón diferente. En

5 algunas realizaciones, se usan dos transposomas diferentes, y los dos transposomas diferentes contienen cada uno la misma transposasa y las composiciones de extremos del transposón contienen diferentes cadenas transferidas. En algunas realizaciones, se usan dos transposomas diferentes, y cada uno de los dos transposomas incluye diferentes enzimas transposasa y diferentes composiciones de extremos de transposón, cada una de las cuales forma un complejo funcional con la respectiva transposasa.

10 En algunas realizaciones, la cantidad de la transposasa y la composición de extremos del transposón o de la composición del transposoma utilizada en la reacción de transposición in vitro está entre aproximadamente 1 picomol y aproximadamente 25 picomoles por 50 nanogramos de ácido nucleico objetivo por reacción de 50 microlitros. En algunas realizaciones, la cantidad de la transposasa y la composición de extremos del transposón o de la composición del transposoma utilizada en la reacción de transposición in vitro es de entre aproximadamente 5 picomoles y aproximadamente 50 picomoles por 50 nanogramos de ácido nucleico objetivo por reacción de 50 microlitros. En algunas realizaciones, la concentración de la transposasa es de 0.5-1 nM. En algunas realizaciones, la concentración de la transposasa es de 0.01-0.02 picomoles por 20 μ l de reacción.

15 El ejemplo 2 ilustra un protocolo para la etapa de tagmentación usando un método proporcionado en el presente documento. En las realizaciones en donde se usa una célula individual para preparar una biblioteca para la secuenciación, solo están presentes dos copias del genoma y, por lo tanto, un tamaño de inserto más pequeño tiende a aumentar la diversidad de la biblioteca. Como se muestra en el Ejemplo 8, los recuentos, y por lo tanto la diversidad representada por una biblioteca, aumentan a medida que disminuye el tamaño del inserto. Por lo tanto, en algunas realizaciones, el método en el presente documento usa una mayor cantidad de transposasa en la etapa de tagmentación para aumentar la fragmentación y reducir el tamaño del inserto de los fragmentos de ácido nucleico marcados. Como se muestra, cuando se usa 1 μ l de Tn5 en una reacción de tagmentación, el tamaño promedio de fragmentos es de aproximadamente 550 pb; mientras que cuando se usan 2 μ l de Tn5 en una reacción de tagmentación, el tamaño promedio de fragmentos es de aproximadamente 400 pb. De acuerdo con un tamaño de inserto más pequeño, la diversidad de la biblioteca aumenta cuando se trata con 2 μ l de Tn5 en comparación con el tratado con 1 μ l de Tn5. Tn5 se usa para ilustrar el ajuste de la transposasa. Debe apreciarse que también se pueden usar otras transposasas en los presentes métodos y su cantidad se puede ajustar y optimizar usando el método proporcionado en el presente documento y los métodos conocidos por los expertos en la técnica.

20 En algunas realizaciones, el tiempo de reacción para la reacción de transposición in vitro es de dos horas o menos, una hora o menos, 30 minutos o menos, 15 minutos o menos, o 10 minutos o menos. En algunas realizaciones, el tiempo de reacción para la reacción de transposición in vitro es de 5 minutos o menos.

25 En algunas realizaciones, la temperatura de reacción para la reacción de transposición in vitro es de aproximadamente 40°C a aproximadamente 70°C, de aproximadamente 45°C a aproximadamente 65°C, o de aproximadamente 50°C a aproximadamente 60°C. En algunas realizaciones, la temperatura de reacción para la reacción de transposición in vitro es de aproximadamente 55°C.

30 En algunas realizaciones, la reacción de transposición in vitro puede terminarse manteniendo la muestra, p. ej., en un tubo, a 4°C. En algunas realizaciones, se neutraliza el tampón de tagmentación a los productos de tagmentación y se incuba la muestra a temperatura ambiente durante 5 minutos.

35 A través de una reacción de transposición in vitro, los fragmentos de ácido nucleico objetivo se marcan en el extremo 5'. En algunas realizaciones, el método proporcionado en el presente documento incluye además etapas para incorporar un marcador en el extremo 3' a los fragmentos de ácido nucleico marcados en 5' para hacer una biblioteca de fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, se genera una biblioteca de fragmentos de ácido nucleico con dos marcadores a partir de ácido nucleico objetivo marcado en 5' en un solo tubo sin realizar ninguna etapa de purificación intermedia. La adición del marcador al extremo 3' se puede realizar mediante varios métodos, p. ej., usando ADN polimerasa, transferasa terminal y/o ligasa como se describe en el documento WO 2010/048605.

40 Por lo tanto, en algunas realizaciones, el método proporcionado en el presente documento comprende además (d) incubar la mezcla de la etapa (c) directamente con al menos una enzima modificadora de ácido nucleico en condiciones en las que un marcador de 3' se une a los fragmentos de ácido nucleico marcados en 5' para generar una pluralidad de fragmentos de ácido nucleico objetivo con dos marcadores. En algunas realizaciones, las etapas (a), (b), (c) y (d) se realizan en un solo tubo de reacción. A continuación, se describen realizaciones que ilustran la generación de una biblioteca de fragmentos de ácido nucleico con dos marcadores.

45 En algunas realizaciones, se generan fragmentos de ácido nucleico con dos marcadores usando una polimerasa, p. ej., una ADN polimerasa, con desplazamiento de cadena o actividad de nucleasa 5'. En algunas realizaciones, el método proporcionado en el presente documento incluye incubar la población de fragmentos de ácido nucleico marcados en 5' reasociados con una ADN polimerasa que tiene desplazamiento de cadena o actividad de nucleasa 5' en condiciones sin termociclado y en donde los fragmentos de ácido nucleico 5' marcados reasociados no están desnaturalizados, en donde la ADN polimerasa extiende el extremo 3' de cada cadena de los fragmentos de ácido nucleico marcados en 5' reasociados usando la cadena complementaria como plantilla y desplaza o digiere la cadena no transferida, generando así la biblioteca de fragmentos de ADN bicatenario con dos marcadores. En una realización,

la etapa de extensión se realizó a 72°C usando el marcador en 5' en la cadena opuesta como plantilla.

En algunas realizaciones, los fragmentos de ADN bicatenario con dos marcadores generados por el método proporcionado anteriormente se desnaturalizan para generar una biblioteca de fragmentos de ADN marcados que contiene fragmentos de ADN bicatenarios con dos marcadores (p. ej., por calentamiento a 95°C y enfriamiento rápido).

5 En otras realizaciones, se generan fragmentos de ácido nucleico con dos marcadores utilizando transferasa terminal. En algunas realizaciones, los fragmentos de ácido nucleico bicatenario marcados en 5' se desnaturalizan para generar los fragmentos de ácido nucleico monocatenario marcados en 5'. Los fragmentos de ácido nucleico monocatenario marcados en 5' se incuban con una ADN polimerasa que consiste en una transferasa terminal y al menos un sustrato para la transferasa terminal durante la cual la transferasa terminal une un segundo marcador en el extremo 3' de fragmentos de ácido nucleico marcados en 5', generando de este modo una biblioteca de fragmentos de ácido nucleico marcados que contienen fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, el extremo 3' del extremo del transposón no transferido que compone la composición de extremos del transposón está bloqueado (p. ej., mediante el uso de un extremo de transposón no transferido que tiene un nucleótido didesoxi o un 3'-O-metil-nucleótido como el nucleótido 3'-terminal), que bloquea el nucleótido 3' y evita la adición por la transferasa terminal, evitando así el marcado de fondo del extremo del transposón no transferido.

10 En otras realizaciones, los fragmentos de ácido nucleico bicatenario marcados en 5' no se desnaturalizan para generar los fragmentos de ácido nucleico monocatenario marcados en 5'. En cambio, los fragmentos de ácido nucleico marcados en 5' se incuban, sin una etapa de desnaturalización previa, con una ADN polimerasa que consiste en una transferasa terminal y al menos un sustrato para la transferasa terminal en condiciones y durante un tiempo suficiente en el que la transferasa terminal une el segundo marcador al extremo 3' de los fragmentos de ácido nucleico marcados en 5', generando así una biblioteca de fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, el extremo 3' del extremo del transposón no transferido que compone la composición de extremos del transposón está bloqueado (p. ej., usando un extremo de transposón no transferido que tiene un didesoxi nucleótido o un 3'-O-metil-nucleótido como el nucleótido 3' terminal).

20 En otras realizaciones, se generan fragmentos de ácido nucleico con dos marcadores usando una ADN polimerasa y un oligonucleótido de marcaje terminal. En algunas realizaciones, los fragmentos de ácido nucleico bicatenario marcados en 5' se desnaturalizan para generar fragmentos de ácido nucleico monocatenario marcados en 5' (p. ej., calentando a 95°C y enfriando rápidamente), y se une un segundo marcador al extremo 3' del fragmento de ácido nucleico monocatenario marcado en 5' usando una ADN polimerasa y un oligonucleótido de marcaje terminal, generando así una biblioteca de fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, las etapas de unir el segundo marcador al extremo 3' de los fragmentos de ácido nucleico marcados en 5' usando una ADN polimerasa y un oligonucleótido de marcaje terminal incluyen: (1) proporcionar un oligonucleótido de marcaje terminal que tiene una porción 5' y porción 3', la porción 5' presenta una secuencia que es complementaria a la secuencia del segundo marcador que se desea unir a los extremos 3' de los fragmentos de ácido nucleico monocatenario marcados en 5', y la porción 3' presenta una secuencia aleatoria que contiene entre tres y ocho nucleótidos aleatorios, de los cuales, el nucleótido 3'-terminal está bloqueado de manera que no puede ser extendido por la ADN polimerasa; (2) poner en contacto los fragmentos de ácido nucleico monocatenario marcados en 5' con el oligonucleótido de marcaje terminal en condiciones y durante un tiempo suficiente en el que el oligonucleótido de marcaje terminal se reasocia con los fragmentos de ácido nucleico monocatenario marcados en 5'; y (3) poner en contacto los fragmentos de ácido nucleico monocatenario marcados en 5' con los que el oligonucleótido de marcaje terminal se reasocia con la ADN polimerasa en una mezcla de reacción y en condiciones de polimerización de ADN y durante un tiempo suficiente en el que los extremos 3' de los fragmentos de ácido nucleico monocatenario marcados en 5' se extienden usando el oligonucleótido de marcaje terminal como plantilla, de modo que el segundo marcador se une a sus extremos 3' y se generan fragmentos de ácido nucleico monocatenario marcados en 5' y 3'.

45 En otras realizaciones más, se generan fragmentos de ácido nucleico con dos marcadores usando una ligasa dependiente de plantilla y un oligonucleótido de marcaje de ligadura. En algunas realizaciones, los fragmentos de ácido nucleico marcados en 5' se incuban con una ADN ligasa dependiente de plantilla y un oligodesoxinucleótido de marcaje de ligadura que tiene una porción 3' y una porción 5', en donde la porción 3' presenta un segundo marcador que presenta cualquier secuencia que se desee unir al extremo 3' de los fragmentos de ADN marcados en 5' y la porción 5' tiene un grupo 5'-monofosfato y presenta una secuencia aleatoria, en condiciones y durante el tiempo suficiente en el que el segundo marcador se une a los fragmentos de ADN marcados en 5' reasociados, generando así una biblioteca de fragmentos de ADN que comprende fragmentos de ADN reasociados con dos marcadores. En algunas realizaciones, el método incluye además la etapa de desnaturalizar la biblioteca de fragmentos de ADN que comprende fragmentos de ADN con dos marcadores reasociados (p. ej., calentando a 95°C y enfriando rápidamente), generando así una biblioteca de fragmentos de ADN monocatenarios dos marcadores.

50 Después de que se genera una biblioteca de fragmentos de ácido nucleico marcados, los fragmentos de ácido nucleico marcados pueden amplificarse, p. ej., utilizando la reacción en cadena de la polimerasa de ciclo limitado (PCR), para introducir otras secuencias terminales o adaptadores, p. ej., índice, cebadores universales y otras secuencias requeridas para la formación y secuenciación de grupos. En algunas realizaciones, dicha amplificación se realiza en una biblioteca de fragmentos de ácido nucleico marcados en 5'. En algunas realizaciones, dicha amplificación se lleva a cabo en una biblioteca de fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, la

amplificación se realiza en el mismo tubo de reacción donde se genera la biblioteca de fragmentos de ácido nucleico marcados, y los agentes para la amplificación se añaden directamente al mismo tubo de reacción.

Por lo tanto, el método proporcionado en el presente documento incluye además (e) amplificar uno o más fragmentos de ácido nucleico objetivo con dos marcadores para generar una biblioteca de fragmentos de ácido nucleico marcados con secuencia adicional en el extremo 5' y/o el extremo 3' de los fragmentos de ácido nucleico con dos marcadores. En algunas realizaciones, las etapas (a), (b), (c), (d) y (e) se realizan en un solo tubo de reacción. Los ejemplos de métodos de amplificación incluyen reacción en cadena de la polimerasa (PCR), reacción de amplificación por desplazamiento de cadena, reacción de amplificación por círculo rodante, reacción en cadena de ligasa, reacción de amplificación mediada por transcripción y reacción de amplificación mediada por bucle.

En algunas realizaciones, el método proporcionado en el presente documento incluye amplificar la biblioteca de fragmentos de ácido nucleico monocatenario con dos marcadores usando una PCR. En algunas realizaciones, el método proporcionado en el presente documento usa la amplificación por PCR con un solo cebador de una biblioteca de fragmentos de ADN con dos marcadores. En algunas realizaciones, la etapa de amplificar fragmentos de ADN con dos marcadores incluye el uso de una ADN polimerasa y al menos un cebador que es complementario al segundo marcador. En algunas realizaciones, la etapa de amplificar la biblioteca de fragmentos de ADN con dos marcadores incluye la amplificación de la biblioteca de fragmentos de ADN marcados por PCR usando solo un oligodesoxirribonucleótido que presenta la secuencia de al menos una porción de la cadena transferida como un cebador de PCR y los fragmentos de ADN con dos marcadores como plantillas. En algunas realizaciones, el cebador contiene una porción 5' que contiene una secuencia adicional, p. ej., una secuencia de adaptador.

En algunas realizaciones, se usan dos cebadores de PCR diferentes, cada uno de cuyos cebadores de PCR muestra la secuencia de al menos una porción del extremo del transposón transferido que compone la composición de extremos del transposón. En algunas realizaciones, cada cebador de PCR incluye una porción 3' y una porción 5', en donde la porción 3' presenta la secuencia de extremo del transposón transferido respectiva y la porción 5' presenta la secuencia de un dominio de marcador respectivo o un adaptador para un propósito particular (p. ej., un dominio de marcador de secuenciación/adaptador o un dominio de marcador de amplificación/adaptador, y opcionalmente un dominio de marcador de dirección/adaptador para la secuenciación o amplificación de próxima generación). Por ejemplo, cuando se usa una sola composición de extremos de transposón en la reacción de transposición in vitro para generar la biblioteca de fragmentos de ADN con dos marcadores usando una ADN polimerasa que tiene desplazamiento de cadena o actividad de nucleasa 5', los fragmentos de ADN con dos marcadores se pueden amplificar por PCR usando dos cebadores de PCR diferentes. Cada cebador de PCR contiene una porción 3' y una porción 5', en donde la porción 3' presenta la secuencia de extremo del transposón transferido respectiva y la porción 5' presenta la secuencia de un dominio de marcador/adaptador respectivo para un propósito particular (p. ej., un dominio de marcador de secuenciación/adaptador o un dominio de marcador de amplificación/adaptador, y opcionalmente un dominio de marcador de dirección/adaptador para la secuenciación o amplificación de próxima generación). En algunas realizaciones, la porción 5' de cada cebador de PCR es diferente de la del otro cebador, y como tal las secuencias de los dos extremos del producto de PCR son diferentes. Por ejemplo, un extremo contiene un índice y/o una secuencia de cebador universal, y el otro extremo contiene un índice y/o una secuencia de cebador universal diferente.

En algunas realizaciones, los dos extremos de los fragmentos de ácido nucleico con dos marcadores se originan a partir de dos secuencias de cadena transferidas diferentes. Por ejemplo, en algunas realizaciones, se pueden usar dos transposomas diferentes en la reacción de transposición in vitro, y cada uno de los dos transposomas contiene la misma transposasa pero una composición de extremos de transposón diferente. En algunas realizaciones, se usan dos transposomas diferentes, y los dos transposomas diferentes contienen cada uno la misma transposasa y las composiciones de extremos de transposón contienen diferentes cadenas transferidas. En algunas realizaciones, se usan dos transposomas diferentes, y cada uno de los dos transposomas incluye diferentes enzimas transposasa y diferentes composiciones de extremos de transposón, cada una de las cuales forma un complejo funcional con la respectiva transposasa. En algunas realizaciones, en donde se usan dos composiciones de extremos de transposón diferentes en la reacción de transposición in vitro, y la biblioteca de fragmentos de ácido nucleico con dos marcadores se genera usando una ADN polimerasa que tiene actividad de desplazamiento de cadena o de nucleasa 5', el primer marcador presenta la secuencia de la cadena transferida de una composición de extremos de transposón y el segundo marcador presenta la secuencia de la cadena no transferida de la otra composición de extremos de transposón.

En las realizaciones mencionadas anteriormente y otras realizaciones en donde se unen dos cadenas transferidas diferentes al extremo 5' de cada cadena opuesta del ácido nucleico bicatenario, el método proporcionado en el presente documento puede incluir además la etapa de amplificar los fragmentos de ácido nucleico con dos marcadores por PCR utilizando dos cebadores de PCR diferentes. Uno de los cebadores de PCR presenta la secuencia de al menos una porción de una cadena transferida que compone una composición de extremos de transposón, y el otro de los cebadores de PCR presenta la secuencia de al menos una porción de la otra cadena transferida que compone la otra composición de extremos de transposón.

En algunas realizaciones en las que se usan dos cebadores, cada cebador de PCR contiene una porción 3' y una porción 5', en donde la porción 3' presenta la secuencia del extremo del transposón transferido respectiva y la porción 5' presenta la secuencia de un dominio de marcador/adaptador respectivo para un propósito particular (p. ej., un dominio de marcador de secuenciación o un dominio de marcador de amplificación, y opcionalmente un dominio de

marcador de dirección para la secuenciación o amplificación de próxima generación). En algunas realizaciones, la porción 5' de cada cebador de PCR es diferente de la del otro cebador, y por ello introducen diferentes secuencias en los dos extremos del producto de PCR. En algunas realizaciones, la porción 5' del primer cebador de PCR o la porción 5' del segundo cebador de PCR, o las porciones 5' tanto del primer como del segundo cebador de PCR contienen primeros o segundos marcadores de secuenciación/adaptadores, respectivamente, para la generación de plantillas para la secuenciación de próxima generación para una plataforma de secuenciación particular (p. ej., marcadores de secuenciación para una plataforma de secuenciación Nextera de Illumina). En algunas realizaciones, la porción 5' del primer cebador de PCR o la porción 5' del segundo cebador de PCR contiene adicionalmente un dominio de marcador de dirección/adaptador u otro dominio de marcador/adaptador para un propósito particular.

El ejemplo 3 ilustra una amplificación por PCR de ciclo limitado que puede añadir otras secuencias en los dos extremos de los fragmentos de ácido nucleico marcados, p. ej., índice 1 (i7) e índice 2 (i5) (de Illumina, Inc, San Diego, CA) y secuencias requeridas para otros fines, p. ej., formación de agrupaciones. En una secuenciación de una sola célula, el ADN de entrada es relativamente pequeño y, por lo tanto, el número de ciclos de PCR se puede ajustar para lograr mejores resultados de secuenciación. En el Ejemplo 9, se ensaya y optimiza el número de ciclos de PCR utilizando una sola célula como material de partida. Como se muestra, el ruido es grande cuando se usa PCR con 16 ciclos en un análisis de número de copias, y el ruido se reduce significativamente cuando se usa PCR con 18 ciclos o 20 ciclos. Por lo tanto, en algunas realizaciones, el número de ciclos de PCR es 18, 19 o 20.

Se encuentra disponible una amplia variedad de enzimas y kits para realizar la reacción de amplificación por PCR como conocen los expertos en la materia. Por ejemplo, en algunas realizaciones, la amplificación por PCR se realiza utilizando el sistema de PCR FAILSAFE™ o el sistema de PCR extralargo MASTERAMP™ de EPICENTER Biotechnologies, Madison, WI, según lo descrito por el fabricante. Sin embargo, la presente descripción no se limita al uso de esos productos o condiciones para la reacción de amplificación y se puede usar cualquier mezcla de reacción y polimerasa de ADN termoestable adecuada que permita la amplificación de la secuencia entre el cebador que se reasocia con la secuencia objetivo y el cebador que se reasocia con el transposón.

El método proporcionado en el presente documento no está limitado al uso de la PCR para amplificar la biblioteca de fragmentos de ácido nucleico marcados. Se puede usar cualquier método de amplificación adecuado (p. ej., amplificación por círculo rodante, amplificación de ribocebadores (p. ej., patente de EE.UU. n° 7,413,857), ICAN, UCAN, Ribospia, marcado terminal (solicitud de patente de EE.UU. n° 20050153333), amplificación de aARN de tipo Eberwine o amplificación por desplazamiento de cadena) que amplifica la misma secuencia y genera una composición y cantidad de producto de amplificación adecuadas para el propósito pretendido, en realizaciones de la presente invención. Por ejemplo, algunos métodos de desplazamiento de cadena que se pueden usar se describen en las publicaciones de patentes PCT n°. WO 02/16639; WO 00/56877; y AU 00/29742; de Takara Shuzo Company, Kyoto, Japón; patentes de EE.UU. n° 5,523,204; 5,536,649; 5,624,825; 5,631,147; 5,648,211; 5,733,752; 5,744,311; 5,756,702; y 5,916,779 de Becton Dickinson and Company; patentes de EE.UU. n° 6,238,868; 6,309,833; y 6,326,173 de Nanogen/Becton Dickinson Partnership; patentes de EE.UU. n° 5,849,547; 5,874,260; y 6,218,151 de Bio Merieux; patentes de EE.UU. n° 5,786,183; 6,087,133; y 6,214,587 de Gen-Probe, Inc.; patente de EE.UU. n° 6,063,604 de Wick et al; patente de EE.UU. n° 6,251,639 de Kurn; patente de EE.UU. n° 6,410,278; y publicación PCT N° WO 00/28082 de Eiken Kagaku Kabushiki Kaishi, Tokio, Japón; patentes de EE.UU. n° 5,591,609; 5,614,389; 5,773,733; 5,834,202; y 6,448,017 de Auerbach; y patentes de EE.UU. n° 6,124,120; y 6,280,949 de Lizardi.

En algunas realizaciones, las bibliotecas de fragmentos de ácido nucleico marcados preparadas por cualquier método de la presente descripción se pueden someter a etapas para purificar el ácido nucleico de la biblioteca y opcionalmente para proporcionar una selección de tamaños. Estas etapas pueden ayudar a limpiar los productos de PCR y eliminar el ácido nucleico con un tamaño indeseable. Se pueden usar varios métodos en la técnica para limpiar fragmentos de ácido nucleico generados en los presentes métodos, que incluyen, pero no se limitan a usar columnas para limpiar los fragmentos, p. ej., usando el kit de purificación de PCR QIAquick de Qiagen, y usando la selección de tamaños en gel, p. ej., utilizando la plataforma de electroforesis Pippin Prep. También se pueden usar otros métodos para limpiar fragmentos de ácido nucleico y/o para seleccionar el tamaño de ácido nucleico conocidos en la técnica en el método proporcionado en el presente documento.

Por ejemplo, en algunas realizaciones, se utilizan perlas AMPure XP (de Beckman Coulter Genomics) para purificar los fragmentos de ácido nucleico marcados. Los fragmentos de ácido nucleico pueden unirse a perlas de inmovilización reversible en fase sólida (SPRI), y la afinidad de los fragmentos de ácido nucleico con diferente longitud por las perlas puede controlarse alterando la concentración de PEG/NaCl. Por lo tanto, al alterar la concentración de PEG/NaCl, se pueden purificar selectivamente ácidos nucleicos con diferentes tamaños. En algunas realizaciones, el método proporcionado en el presente documento usa un único tratamiento de AMPure XP para eliminar fragmentos de ácido nucleico por debajo de un cierto tamaño (p. ej., 150-200 pb). En algunas realizaciones, se puede realizar una selección de tamaño doble (superior e inferior) mediante dos etapas consecutivas de AMPure XP. En la primera etapa de selección, se añade una concentración baja de perlas AMPure XP a la muestra para unir fragmentos de ADN más grandes. En esta etapa, se descartan las perlas que contienen los fragmentos más grandes. Luego, en la segunda etapa de selección, se añaden más perlas al líquido sobrenadante. En esta segunda etapa, la cantidad de PEG y NaCl aumenta para que los fragmentos de menor tamaño se unan. A continuación, el líquido sobrenadante que contiene fragmentos de biblioteca muy cortos se descarta y las perlas se lavan y se eluyen los fragmentos intermedios. Los expertos en la técnica entenderán que, dependiendo de las concentraciones de PEG y NaCl en la primera y última

etapa de SPRI, se pueden generar distintos intervalos de tamaños como se ilustra en Bronner et al., 2009, *Curr Protoc Hum Genet.* 18:10.

El procedimiento típico para limpiar una biblioteca de fragmentos de ácido nucleico usando perlas AMPure XP incluye (1) agitar con vórtice las perlas AMPure XP para asegurar que las perlas se dispersen uniformemente; (2) añadir cierta cantidad de perlas AMPure XP a cada producto de PCR generado e incubar a temperatura ambiente; (3) colocar los tubos en un soporte para tubos en el soporte magnético hasta que el líquido sobrenadante se haya aclarado; (5) retirar y descartar el líquido sobrenadante; (6) sin quitar los tubos del soporte magnético, lavar las perlas una o varias veces; (7) con los tubos todavía en el soporte magnético, permitir que las perlas se sequen al aire; (8) retirar los tubos del soporte magnético y añadir tampón de resuspensión e incubar a temperatura ambiente; y (9) transferir el líquido sobrenadante a tubos nuevos.

Después de limpiar la biblioteca de fragmentos de ácido nucleico y seleccionar el tamaño, se puede someter a una etapa de normalización de la biblioteca para normalizar la cantidad de cada biblioteca y garantizar una representación de biblioteca aproximadamente igual en cada muestra agrupada. En algunas realizaciones, se usa un procedimiento de normalización de biblioteca basado en perlas en el método proporcionado en el presente documento. En un procedimiento de normalización de la biblioteca basado en perlas, se añade una cantidad aproximadamente igual de perlas a cada pocillo que contiene una muestra de fragmentos de ácido nucleico. Debido a que la cantidad de perlas añadidas en cada pocillo es aproximadamente igual, la cantidad de fragmentos de ácido nucleico unidos a las perlas también es aproximadamente igual en cada pocillo. Por ello, después de retirar el líquido sobrenadante, y eluir los fragmentos de ácido nucleico de las perlas pueden estar en una cantidad aproximadamente igual en cada pocillo.

Un procedimiento típico de normalización de la biblioteca basada en perlas incluye (1) añadir una cantidad aproximadamente igual de perlas (p. ej., en un tampón de perlas) en cada pocillo que contiene fragmentos de ácido nucleico generados en los métodos proporcionados anteriormente; (2) incubar y/o agitar para permitir la unión de las perlas con fragmentos de ácido nucleico; (3) colocar pocillos (pueden estar en una placa) en un soporte magnético y dejar que el líquido sobrenadante se aclare; (4) con los pocillos en el soporte magnético, retirar y descartar cuidadosamente el líquido sobrenadante; (5) lavar perlas una o varias veces; y (6) eluir los fragmentos de ácido nucleico unidos a las perlas.

En algunas realizaciones, la biblioteca de fragmentos de ácido nucleico marcados generados por el método proporcionado en el presente documento puede usarse como plantillas para la secuenciación de ácidos nucleicos.

En algunas realizaciones, antes de la secuenciación, los fragmentos de ácido nucleico marcados en la biblioteca se amplifican para intensificar las señales frente al ruido durante una secuenciación. p. ej., en una secuenciación por síntesis. En algunas realizaciones, la biblioteca de fragmentos de ácido nucleico marcados se usa como plantilla para una reacción de amplificación (p. ej., una reacción de amplificación por PCR usando cebadores de PCR que son complementarios a las secuencias de extremos de los fragmentos de ácido nucleico marcados). En algunas realizaciones, la biblioteca de fragmentos de ácido nucleico marcados amplificados contiene la mayoría o aproximadamente todas las secuencias presentadas por el ácido nucleico objetivo. En algunas realizaciones en donde el ácido nucleico objetivo incluye el ADN genómico de un organismo, la reacción de amplificación es una reacción de amplificación del genoma completo.

En algunas realizaciones, los fragmentos de ácido nucleico marcados pueden inmovilizarse sobre una superficie sólida. Por ejemplo, la superficie sólida se puede unir con un polinucleótido complementario a una secuencia de extremo de fragmentos de ácido nucleico marcados, y por ello los fragmentos de ácido nucleico marcados se pueden inmovilizar en la superficie sólida. Luego, los fragmentos de ácido nucleico inmovilizados se amplifican en la superficie. Por ejemplo, en algunas realizaciones, los fragmentos de ácido nucleico inmovilizados se amplifican usando metodologías de amplificación en grupos como se ilustra en las descripciones de las patentes de EE.UU. nº 7,985,565 y 7,115,400. Los materiales de las patentes de EE.UU. nº 7,985,565 y 7,115,400 describen métodos de amplificación de ácidos nucleicos en fase sólida que permiten que los productos de amplificación se inmovilicen sobre un soporte sólido con el fin de formar matrices compuestas de grupos o "colonias" de moléculas de ácido nucleico inmovilizadas. Cada grupo o colonia en dicha matriz se forma a partir de una pluralidad de cadenas polinucleotídicas inmovilizadas idénticas y una pluralidad de cadenas polinucleotídicas complementarias inmovilizadas idénticas. Las matrices así formadas generalmente se denominan en el presente documento "matrices agrupadas". Los productos de las reacciones de amplificación en fase sólida como los descritos en las patentes de EE.UU. nº 7,985,565 y 7,115,400 son las denominadas estructuras "con puente" formadas por la reasociación de pares de cadenas de polinucleótidos inmovilizadas y cadenas complementarias inmovilizadas, estando ambas cadenas inmovilizadas en el soporte sólido en el extremo 5', p. ej., mediante una unión covalente. Las metodologías de amplificación en grupos son ejemplos de métodos en los que se usa una plantilla de ácido nucleico inmovilizado para producir amplicones inmovilizados. Otras metodologías adecuadas conocidas en la técnica también se pueden usar para producir amplicones inmovilizados a partir de fragmentos de ácido nucleico marcados inmovilizados producidos según los métodos proporcionados en el presente documento.

La biblioteca de fragmentos de ácido nucleico marcados preparados de acuerdo con el método proporcionado en el presente documento puede secuenciarse de acuerdo con cualquier metodología de secuenciación adecuada, tal como secuenciación directa, incluyendo secuenciación por síntesis, secuenciación por ligadura, secuenciación por

hibridación, secuenciación por nanoporos y similares. En algunas realizaciones, los fragmentos de ADN inmovilizados se secuencian sobre un soporte sólido. En algunas realizaciones, el soporte sólido para la secuenciación es el mismo soporte sólido sobre el que se produce la amplificación.

5 En algunas realizaciones, la metodología de secuenciación utilizada en el método proporcionado en el presente documento es la secuenciación por síntesis (SBS). En la SBS, la extensión de un cebador de ácido nucleico a lo largo de una plantilla de ácido nucleico (p. ej. un ácido nucleico objetivo o un amplicón del mismo) se controla para determinar la secuencia de nucleótidos en la plantilla. El proceso químico subyacente puede ser la polimerización (p. ej. catalizada por una enzima polimerasa). En una realización particular de SBS basada en polimerasa, los nucleótidos marcados con fluorescencia se añaden a un cebador (extendiendo así el cebador) de una manera dependiente de la
10 plantilla, de modo que la detección del orden y el tipo de nucleótidos añadidos al cebador se puede usar para determinar la secuencia de la plantilla.

Se pueden usar otros procedimientos de secuenciación que usan reacciones cíclicas, tales como la pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) a medida que se incorporan nucleótidos particulares en una cadena nascente de ácido nucleico (Ronaghi, et al., 1996, *Analytical Biochemistry* 242 (1), 84-9; Ronaghi, 2001, *Genome Res.* 11 (1), 3-11; Ronaghi et al., 1998, *Science* 281 (5375), 363; US 6,210,891; US 6,258,568 y US 6,274,320). En la pirosecuenciación, el PPi liberado puede detectarse al convertirse inmediatamente en trifosfato de adenosina (ATP) por la ATP sulfurilasa, y el nivel de ATP generado puede detectarse a través de fotones producidos por la luciferasa. Por lo tanto, la reacción de secuenciación se puede controlar a través de un sistema de detección de
15 luminiscencia. Las fuentes de radiación de excitación utilizadas para los sistemas de detección basados en fluorescencia no son necesarias para los procedimientos de pirosecuenciación. Los sistemas fluidicos, detectores y procedimientos útiles que pueden adaptarse para la aplicación de pirosecuenciación a amplicones producidos de acuerdo con la presente descripción se describen, por ejemplo, en los documentos US 2005/0191698 A1, US 7,595,883 y US 7,244,559.

Algunas realizaciones pueden utilizar métodos que implican el seguimiento en tiempo real de la actividad de la ADN polimerasa. Por ejemplo, las incorporaciones de nucleótidos se pueden detectar a través de interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa que lleva fluoróforo y nucleótidos marcados con γ -fosfato, o con guías de ondas en modo cero (ZMW). Las técnicas y reactivos para la secuenciación basada en FRET se describen, por ejemplo, en Levene et al., 2003, *Science* 299, 682-686; Lundquist et al., 2008, *Opt. Lett.* 33, 1026-1028; Koriach et al., 2008, *Proc. Natl. Acad. Sci. USA* 105, 1176-1181.
25

Algunas realizaciones de SBS incluyen la detección de un protón liberado tras la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede usar un detector eléctrico y técnicas asociadas que están disponibles comercialmente en Ion Torrent (Guilford, CT, una subsidiaria de Life Technologies) o los métodos y sistemas de secuenciación descritos en los documentos US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; o US 2010/0282617 A1. Los métodos expuestos en el presente documento para amplificar ácidos nucleicos objetivo usando exclusión cinética se pueden aplicar fácilmente a sustratos utilizados para detectar protones. Más específicamente, los métodos expuestos en el presente documento pueden usarse para producir poblaciones clonales de amplicones que se usan para detectar protones.
30

Otra técnica útil de secuenciación es la secuenciación de nanoporos (véase, por ejemplo, Deamer et al., 2000, *Trends Biotechnol.*, 18, 147-151; Deamer et al., 2002, *Acc. Chem Res.* 35: 817-825; Li et al., 2003, *Nat. Mater.* 2: 611-615). En algunas realizaciones de nanoporos, el ácido nucleico objetivo o los nucleótidos individuales eliminados de un ácido nucleico objetivo pasan a través de un nanoporo. A medida que el ácido nucleico o nucleótido pasa a través del nanoporo, se puede identificar cada tipo de nucleótido midiendo las fluctuaciones en la conductancia eléctrica del poro. (patente de EE.UU. nº 7,001,792; Soni et al., 2007, *Clin. Chem.*, 53, 1996-200; Healy, 2007, *Nanomed.* 2, 459-481; Cockroft et al., 2008, *J. Am. Chem Soc.*, 130, 818-820).
40

En algunas realizaciones, el método proporcionado en el presente documento incluye además analizar la variación del número de copias de una célula. Un análisis del número de copias prueba la variación del número de copias de ADN en una muestra. Dicho análisis ayuda a detectar la variación del número de copias cromosómicas que puede causar o aumentar los riesgos de varios trastornos críticos. Por ejemplo, se ha descrito que el autismo está asociado con mutaciones en el número de copias (Sebat et al., 2007, Strong association of de novo copy number mutations with autism, *Science* 316 (5823): 445-9). También se ha descrito que la esquizofrenia está asociada con variaciones en el número de copias (St Clair, 2008, Copy number variation and schizophrenia, *Schizophr Bull* 35 (1): 9-12). Se han desarrollado varios métodos para detectar la variación del número de copias. Sin embargo, cuando el material de partida es limitado y proviene de una población mínima de células, el ruido es significativo y el resultado se ve comprometido. El presente método proporciona un método para detectar la variación del número de copias en dicha
45 situación. Los ejemplos proporcionados a continuación demuestran el análisis de la variación del número de copias usando los métodos presentes y varios parámetros están optimizados para el análisis de la variación del número de copias. En algunas realizaciones, la población mínima de células utilizada en el análisis de la variación del número de copias contiene una, dos, tres, cuatro o cinco células. Típicamente, a medida que aumenta el número de células, se puede lograr una distribución de lecturas más completa y, por lo tanto, hay menos ruido presente en los datos como se muestra en el Ejemplo 10. En este ejemplo, la distribución de lecturas usando una, tres o cinco células analizada en este ejemplo. Como se muestra, la cobertura genómica aumenta a medida que aumenta el número de células, se
50
55
60

estima que una célula puede cubrir aproximadamente 40% del genoma, y tres células pueden cubrir más de 50% del genoma, y cinco células pueden cubrir aproximadamente 60% del genoma. Los recuentos promedios de la biblioteca con una célula, tres células y cinco células es de aproximadamente 5 millones, 15 millones y 20 millones, respectivamente. También se muestra en este ejemplo, cuando se usa una sola célula, la tasa de éxito general es relativamente alta 94% (N = 187). Los fallos en el ensayo de una célula probablemente son causados por la calidad de la célula misma, p. ej., selección de una de las células replicantes o células apoptóticas/necróticas.

El ejemplo 11 compara el presente método con algunos métodos actuales de preparación de células individuales. Cuando se usa el kit REPLI-g de célula única desarrollado por QIAGEN (San Diego, CA) para la preparación de ácido nucleico, los datos de variación del número de copias tienen mucho ruido cuando derivan de una sola célula, tres células o cinco células. Cuando se usa SurePlex (PicoPlex) desarrollado por Illumina, Inc (San Diego, CA) para preparar el ácido nucleico, se reduce el ruido en comparación con el kit de células individuales REPLI-g. Como se muestra, el método presente (Nextera SC) reduce aún más el ruido en comparación con el uso del sistema de amplificación SurePlex. Por lo tanto, el presente método proporciona un método avanzado para analizar la variación del número de copias.

Un aspecto del análisis de la variación del número de copias es detectar mosaicismos. Un mosaico o mosaicismos denota la presencia de dos o más genotipos en un individuo. Hay dos tipos principales de mosaicismos: el mosaicismos somático y el mosaicismos de la línea germinal. El mosaicismos somático ocurre cuando las células somáticas contienen más de un genotipo, p. ej., debido a errores mitóticos en la primera o posterior escisiones. Los investigadores han mostrado que las mutaciones somáticas están cada vez más presentes a lo largo de la vida y son responsables de muchas leucemias, linfomas y tumores sólidos (Jacobs et al., 2012, Detectable Clonal Mosaicism and Its Relationship to Aging and Cancer, *Nature Genetics* 44 (6): 651-U668). En el mosaicismos de la línea germinal, algunos gametos (espermatozoides u ovocitos) portan una mutación, pero el resto son normales, lo que también conduce a muchas enfermedades. Por lo tanto, la detección de mosaicismos puede proporcionar información diagnóstica valiosa. La presente descripción proporciona métodos para detectar mosaicismos. En el Ejemplo 12, se ilustra el uso del método proporcionado en el presente documento para detectar mosaicismos. Como se muestra, se detecta una población que representa ADN de 15.4 MB en cada secuenciación de célula individual en un análisis de número de copias del cromosoma 18 de una sola célula GM50121. De manera similar, los datos de análisis del número de copias de los cromosomas 15, X y 10 usando una sola GM20916, y los datos de análisis del número de copias de los cromosomas 1 y 11 usando una sola célula GM10239 detectan ambas poblaciones adicionales que representan otros cromosomas.

Los presentes métodos también se pueden usar para otras aplicaciones, p. ej., cribado genético previo a la implantación, investigación de células individuales, análisis de células tumorales circulantes, biopsia por aspiración con aguja fina, capa leucocitaria y análisis de amniocitos. En estas aplicaciones, el material de ácido nucleico con el que se empieza suele ser limitado y, por lo tanto, el presente método puede mejorar el análisis para estas aplicaciones. Además del análisis de variación del número de copias, el presente método también se puede utilizar para detectar la variante de un solo nucleótido presente en una población mínima de células en las aplicaciones mencionadas anteriormente. La variante de un solo nucleótido incluye polimorfismo de un solo nucleótido (SNP) y mutación puntual. El polimorfismo de un solo nucleótido (SNP) es un tipo común de variación genética que incluye el polimorfismo en una posición del ADN en la que ocurren dos o más bases alternativas con una frecuencia apreciable en la población de personas (generalmente más o igual a 1%). Las mutaciones puntuales son variaciones de base con una frecuencia inferior a 1%. El polimorfismo de un solo nucleótido (SNP) y las mutaciones puntuales representan la mayor fuente de diversidad en el genoma de un ser humano. Estos polimorfismos de un solo nucleótido (SNP) y mutaciones puntuales pueden servir como marcadores biológicos para localizar una enfermedad en el mapa del genoma humano porque generalmente se encuentran cerca de un gen asociado con una determinada enfermedad. Por lo tanto, la detección de polimorfismos de un solo nucleótido (SNP), mutaciones puntuales y mutaciones similares son de gran importancia para actividades clínicas, la salud humana y el control de enfermedades genéticas. El presente método proporciona la ventaja de un acceso uniforme al ADN genómico y ayuda a preservar el material de ácido nucleico objetivo. Por lo tanto, puede mejorar la detección de variación de un solo nucleótido utilizando una población mínima de células.

En la descripción de algunas realizaciones de los diversos métodos anteriores, se usa "tubo de reacción" o "tubo". Debe apreciarse que también se pueden usar otros medios y/o recipientes de reacción en los presentes métodos.

Kits para preparar una biblioteca de fragmentos de ácido nucleico marcados (no forma parte de la invención)

En otro aspecto, la presente descripción proporciona un kit para preparar una biblioteca de fragmentos de ácido nucleico marcados que comprende: (a) un reactivo de lisis que tiene una o más proteasas, y (b) una composición de reacción de transposición que tiene al menos una transposasa y al menos una composición de extremos de transposón que contiene una cadena transferida.

En algunas realizaciones, el reactivo de lisis proporcionado incluye solo una proteasa que tiene una amplia especificidad y, por lo tanto, las proteasas pueden digerir diversas proteínas y polipéptidos. En algunas otras realizaciones, el reactivo de lisis proporcionado en el presente documento incluye una mezcla de diversas proteasas, y la combinación de diversas proteasas puede digerir diversas proteínas y polipéptidos. Las proteasas de ejemplo proporcionadas en el presente documento incluyen serina proteasas, treonina proteasas, cisteína proteasas, aspartato proteasas, ácido glutámico proteasas y metaloproteasas. La proteasa de ejemplo usada en el presente documento

incluye una serina proteasa aislada de una cepa de *Bacillus* recombinante. Las proteasas de ejemplo usadas en el presente documento incluyen subtilisina y variantes de la misma, que incluyen subtilisina Carlsberg, alcalasa y subtilisina S41. Los expertos en la técnica conocen las subtilisinas y variantes de las mismas e incluyen, por ejemplo, alcalasa, alcalasa 0.6L, alcalasa 2.5L, enzima ALK, bacilopeptidasa A, bacilopeptidasa B, bioprasa proteínasa alcalina de *Bacillus subtilis*, bioprasa AL 15, bioprasa APL 30, colistinasa, subtilisina J, subtilisina S41, subtilisina Sendai, subtilisina GX, subtilisina E, subtilisina BL, genenasa I, esperasa, maxatasa, termoasa PC 10, proteasa XXVII, termoasa, superasa, subtilisina Carlsberg subtilisina DY, subtilopeptidasa, SP 266, savinasa 8.0L, savinasa 4.0T, kazusasa, proteasa VIII, opticlean, protina A 3L, savinasa, savinasa 16.0L, savinasa 32.0 L EX, orientasa 10B, proteasa S, serina endopeptidasa. En realizaciones particulares de los métodos y composiciones presentados en el presente documento, se puede usar una proteasa termolábil tal como la subtilisina y las variantes termolábiles de subtilisina, tal como se representa mediante la descripción de ejemplo de Davail et al., 1994, *J. Biol. Chem.*, 26: 17448-17453.

En algunas realizaciones, el reactivo de lisis incluye uno o más detergentes. En algunas realizaciones, el detergente proporcionado en el presente documento no interfiere con las actividades enzimáticas corriente abajo. Por lo tanto, en algunas realizaciones, el reactivo de lisis incluye detergentes no iónicos. Típicamente, los detergentes no iónicos contienen grupos de cabeza hidrófilos sin carga. Los detergentes no iónicos típicos se basan en polioxietileno o un glucósido. Ejemplos de detergentes no iónicos incluyen Tween® 80, Tween® 20, Triton® X-100, Triton® X-100-R, Triton® X-114, NP-40, Genapol® C-100, Genapol® X-100, Igepal® CA 630, Arlasolve® 200, series Brij® 96/97, Triton, Brij® 98, Brij® 58, Brij® 35, Pluronic® L64, Pluronic® P84, sulfobetainas no detergentes (NDSB 201), anfipoles (PMAL-C8), CHAPS, octil β-D-glucopiranosido, saponina, monododecil éter de nonaetilenglicol (C12E9, polidoceno), dodecilsulfato de sodio, N-laurilsarcosina, desoxicolato de sodio, sales biliares, bromuro de hexadeciltrimetilamonio, SB3-10, SB3-12, amidosulfobetaina-14, octil tioglucósido, maltósidos, HEGA y series MEGA. En una realización, el reactivo de lisis incluye componentes proporcionados en las Tablas 1-3.

En algunas realizaciones, la composición de transposición contiene al menos una transposasa y al menos una composición de extremos de transposón que incluye (i) una cadena transferida que tiene una porción 3' que presenta la secuencia de extremo de transposón transferida y una porción 5' que presenta la secuencia para un dominio de marcador para usar en una secuenciación de próxima generación o reacción de amplificación, y (ii) una cadena no transferida que contiene fosfato 5' que presenta solo la secuencia de extremo del transposón no transferido, en donde la transposasa forma un complejo con la composición de extremos de transposón que está activa en una reacción de transposición *in vitro*. En algunas realizaciones, el kit incluye además un tampón de reacción que contiene dimetilformamida en una cantidad que da como resultado que esté presente en la reacción de transposición *in vitro* en una concentración final de 10%. En algunas realizaciones, el dominio de marcador incluye uno o más de un dominio de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección y un dominio de marcador de dirección.

En algunas realizaciones, la composición de la reacción de transposición incluye dos o más composiciones de extremos de transposón, cada una de las dos o más composiciones de extremos de transposón incluye una cadena transferida que difiere en al menos un nucleótido.

En algunas realizaciones, la transposasa es una transposasa de Tn5. En algunas realizaciones, la composición de extremos de transposón incluye un extremo del transposón Tn5. En una realización del kit, el transposoma incluye una transposasa Tn5 o transposasa MuA de tipo natural o hiperactiva que se proporciona en una concentración en la que la concentración final del transposoma en la reacción de transposición *in vitro* es de al menos 250 nM. En algunas otras realizaciones, las concentraciones finales de transposoma Tn5 o transposoma MuA de tipo natural o hiperactivo es al menos 500 nM.

En una realización, la transposasa en el kit es una forma de tipo natural o mutante de la transposasa Tn5 (p. ej., transposasa EZ-Tn5™) en una concentración mayor o igual a aproximadamente 5 unidades por microlitro; aproximadamente 10-20 unidades por microlitro; aproximadamente 20-40 unidades por microlitro; aproximadamente 40-60 unidades por microlitro; aproximadamente 60-80 unidades por microlitro; o aproximadamente 80-100 unidades por microlitro. En algunas realizaciones, el kit proporcionado en el presente documento incluye componentes proporcionados en la Tabla 6.

En algunas realizaciones, el kit adicional incluye una enzima modificadora. En algunas realizaciones, la enzima modificadora es una polimerasa o una ligasa. En algunas realizaciones, el kit incluye al menos otro componente enzimático seleccionado entre: una ADN polimerasa que tiene actividad de nucleasa 5' o actividad de desplazamiento de cadena; una ADN polimerasa que carece de actividad de nucleasa 5', una NAD ligasa dependiente de plantilla y una ligasa independiente de plantilla. En algunas realizaciones, el al menos otro componente enzimático se selecciona entre: mezcla de ADN polimerasa FAILSAFE™; ADN polimerasa Taq, ADN polimerasa Tfi, ADN polimerasa T4, ADN ligasa de *E. coli*, ARN ligasa termoestable del bacteriófago TS2126, ARN ligasa termoestable Mth Rn 1 y ADNmc ligasa termoestable CIRCLIGASE™.

En algunas realizaciones en donde la al menos una enzima en el kit es una ligasa dependiente de plantilla (p. ej., ADN ligasa de *E. coli*), una alta proporción de las moléculas de ligasa están adeniladas y no se proporciona ATP en el kit. En algunas realizaciones en donde la al menos una enzima en el kit es una ligasa dependiente de plantilla (p. ej., ADN ligasa de *E. coli*), el kit incluye adicionalmente un oligonucleótido de marcaje de ligadura que comprende una porción

3' y una porción 5', en donde la porción 3' presenta una secuencia de un dominio de marcador y la porción 5' presenta una secuencia aleatoria que consiste en aproximadamente tres a aproximadamente ocho nucleótidos. En algunas realizaciones, el oligonucleótido de marcaje de ligadura incluye una porción 5' que presenta una secuencia aleatoria que consta de cuatro nucleótidos.

5 En algunas realizaciones en donde la al menos una enzima en el kit es una ligasa independiente de plantilla, seleccionada de entre ARN ligasa termoestable del bacteriófago TS2126, ARN ligasa termoestable Mth Rn 1 y ADNmc ligasa termoestable CIRCLIGASE™, la ligasa independiente de plantilla se proporciona en una forma altamente adenilada y no se proporciona ATP en el kit. En una realización del kit incluye la transposasa EZ-Tn5™ y la ácido nucleico ligasa independiente de plantilla, la composición de extremos de transposón EZ-Tn5 pMEDS incluye tanto
10 una cadena transferida de EZ-Tn5 METS que tiene un grupo 5' -monofosfato como una cadena no transferida de EZ-Tn5 pMENTS que tiene un grupo 5'-monofosfato.

En algunas realizaciones, el kit incluye además un reactivo para una reacción de amplificación. En algunas realizaciones, el reactivo para la reacción de amplificación es un reactivo para PCR. En algunas realizaciones, el reactivo para la reacción de amplificación incluye al menos un cebador. En algunas realizaciones, el al menos un
15 cebador incluye una porción 3' que presenta la secuencia de al menos una porción de la cadena transferida. En algunas realizaciones, el al menos un cebador incluye una porción 5' que contiene una secuencia universal.

En algunas realizaciones, el kit incluye dos cebadores, cada cebador de PCR contiene una porción 3' y una porción 5', en donde la porción 3' presenta la secuencia de extremo del transposón transferido respectiva y la porción 5' presenta la secuencia de un dominio de marcador/adaptador respectivo para un propósito particular (p. ej., un dominio de marcador de secuenciación o un dominio de marcador de amplificación, y opcionalmente un dominio de marcador de dirección para la secuenciación de próxima generación o amplificación). En algunas realizaciones, la porción 5' de
20 cada cebador de PCR es diferente de la del otro cebador. En algunas realizaciones, la porción 5' del primer cebador de PCR o la porción 5' del segundo cebador de PCR, o las porciones 5' tanto del primer como del segundo cebador de PCR contienen marcadores/adaptadores de secuenciación, respectivamente. En una realización, el kit proporcionado en el presente documento incluye los componentes proporcionados en la Tabla 7.

En algunas realizaciones, el kit incluye además un reactivo de selección de tamaño. En algunas realizaciones, el reactivo de selección de tamaño incluye perlas AMPure XP (de Beckman Coulter Genomics). Los fragmentos de ácido nucleico pueden unirse a perlas de inmovilización reversible en fase sólida (SPRI). En algunas realizaciones, el reactivo de selección de tamaño incluye además PEG y NaCl.

30 En algunas realizaciones, el kit proporcionado en el presente documento incluye además un reactivo de normalización de biblioteca. En algunas realizaciones, el reactivo de normalización de la biblioteca incluye los aditivos de normalización de la biblioteca proporcionados por Illumina, Inc (San Diego, CA, componente nº 15025391) y las perlas de normalización de la biblioteca proporcionados por Illumina, Inc (componente nº 15022566). En algunas realizaciones, el reactivo de normalización de la biblioteca incluye además el lavado de normalización de la biblioteca proporcionado por Illumina, Inc (componente nº 15022565). En algunas realizaciones, el reactivo de normalización de la biblioteca incluye además el tampón de almacenamiento de normalización de la biblioteca proporcionado por Illumina, Inc (San Diego, CA, componente nº 15025139).

En algunas realizaciones, el kit incluye además un aparato que tiene una superficie sólida. En algunas realizaciones, a la superficie sólida está unida una población de oligonucleótidos. En algunas realizaciones, el aparato es un aparato de celda de flujo. En algunas realizaciones, la superficie sólida incluye una superficie grabada adecuada para la inmovilización de moléculas en un patrón ordenado.

Los siguientes ejemplos se proporcionan a modo de ilustración, no de limitación.

Ejemplos

Ejemplo 1 Generación de lisado celular que contiene ácido nucleico objetivo

45 En algunas realizaciones, durante la etapa de generación de un lisado celular, las membranas celulares se rompen por el detergente durante lo cual se rompe las asociaciones proteína-lípido y lípido-lípido, y se liberan así materiales intracelulares en forma soluble. La función principal de la proteasa de especificidad amplia es eliminar las proteínas de unión al ADN, tales como las histonas del ADN para permitir el acceso uniforme de la transposasa al ADN. En algunas realizaciones, como se ilustra en este ejemplo, el detergente y la proteasa proporcionados están en una única
50 mezcla de reactivo de lisis. La mezcla se aplica directamente a las células para generar un lisado celular que contiene el ácido nucleico objetivo. Como se discutió anteriormente, en algunas realizaciones, cuando se usa calor para inactivar la proteasa, es importante que el calor no desnaturalice el ácido nucleico bicatenario y que se asegure que no se interfiere en la etapa de tagmentación.

En este ejemplo, la proteasa se puede inactivar por calor a 70°C, y a esta temperatura se preserva la conformación bicatenaria del ADN. En el ejemplo 1 se ilustra un protocolo para la generación de un lisado celular que contiene ácido nucleico objetivo de la siguiente manera:

55

(1) Se mezclan adecuadamente los reactivos invirtiendo suavemente y dando golpecitos a los tubos 3-5 veces, seguido de una breve centrifugación en una microcentrífuga.

5 (2) En un tubo limpio de microcentrífuga, se combinan los componentes de la Tabla 1 para hacer la mezcla principal de lisis 5X. La mezcla principal de lisis se puede ampliar de acuerdo con el número de muestras, p. ej., se puede incluir un 10% adicional para compensar pérdidas durante el pipeteo.

Tabla 1 Componentes de la mezcla principal de lisis

Componente de la mezcla principal de lisis	Volumen (µl)
5X tampón de lisis	1.1
5X Solución madre de proteasa X	1.1
Total	2.2

El tampón de lisis 5X en la Tabla 1 anterior se puede preparar de acuerdo con la siguiente Tabla 2:

Tabla 2 Componentes del tampón de lisis 5X

Componente	Concentración solución madre	Concentración mezcla principal 5x	Volumen (µl)
Tris-HCl (pH 8.0)	1 M	250 mM	250
EDTA	0.5 M	5 mM	10
Triton X-100	10 %	2.5 %	250
Super Q H ₂ O			490
Total			1000

10 Todos los reactivos se pueden mezclar adecuadamente agitando suavemente con vórtice el tubo varias veces, seguido de una breve centrifugación en una microcentrífuga. Esta etapa puede repetirse 3-5 veces. El tampón de lisis 5X se puede almacenar a temperatura ambiente para evitar la precipitación del detergente.

15 La solución madre de proteasa 5X se puede preparar de la siguiente manera: (i) preparar partes alícuotas de almacenamiento de un solo uso resuspendiendo una proteasa, p. ej., la proteasa QIAGEN, directamente en el vial de vidrio añadiendo 2.38 ml de Super Q H₂O a una concentración final de 3150 mUA/ml. Asegurarse de que la proteasa se disuelva adecuadamente agitando suavemente con vórtice el vial varias veces. Repartir la solución en partes alícuotas de 25 µl y congelar inmediatamente a -80°C, y (ii) retirar una parte alícuota de almacenamiento de un solo uso del congelador y descongelar y preparar la solución madre de proteasa 5X de acuerdo con la Tabla 3 a continuación:

20 Tabla 3 Componentes de la solución madre de proteasa 5X

Componente	Concentración madre	Concentración de la mezcla principal 5X	Volumen (µl)
Proteasa QIAGEN	3150 mUA/ml	450 mUA/ml	15
Super Q H ₂ O			90
Total			105

En consecuencia, la concentración final de la solución madre de proteasa 5X es de 450 mUA/ml.

25 (3) Añadir 2 µl de la mezcla principal de lisis preparada anteriormente a cada tubo que contiene una célula, ADN genómico de control positivo o el control negativo. Incubar las muestras de acuerdo con el siguiente programa en un termociclador: 50°C 30 min, 70°C 20 min y 4°C.

En algunas realizaciones, se incluye un ADN genómico de control positivo (aproximadamente 30 pg) en cada experimento. Se puede preparar un ADN genómico de control positivo en una dilución en serie de dos pasos a partir de una solución madre de 10 ng/µl como se prepara en las Tablas 4 y 5 a continuación:

30

Tabla 4 Componente de dilución intermedia de ADN genómico

Componente	Concentración madre	Concentración intermedia	Volumen (µl)
ADN	10 ng/µl	100 pg/µl	2
1X RS1			198
Total			200

Luego, la dilución de ADN intermedia preparada de acuerdo con la tabla anterior se puede diluir posteriormente de acuerdo con la siguiente Tabla 5:

5 Tabla 5 Componente de dilución final de ADN genómico

Componente	Concentración madre	Concentración intermedia	Volumen (µl)
ADN	100 pg/µl	10 pg/µl	10
1X PBS			90
Total			100

Se pueden usar 3 µl de la dilución final preparada en la tabla anterior como entrada de un ADN genómico de control positivo. Esto corresponde a 30 pg o el equivalente genómico de 5 células. También se puede usar más o menos ADN genómico de acuerdo con el método proporcionado en el presente documento.

10 Ejemplo 2 Tagmentación de ácido nucleico objetivo directamente en el lisado celular

En algunas realizaciones, el ADN genómico en el lisado celular, p. ej., como se preparó en el Ejemplo 1, se puede tagmentar (marcar y fragmentar) mediante el transposoma Nextera (disponible en Illumina, Inc, San Diego, CA). El transposoma Nextera puede fragmentar simultáneamente el ADN de entrada y añade secuencias de marcador/adaptador a los extremos. La mezcla principal de tagmentación se puede añadir directamente al lisado celular preparado en el Ejemplo 1 sin ninguna etapa previa de purificación o amplificación de ADN. La mezcla principal de tagmentación se puede preparar como se muestra en la Tabla 6 a continuación y la mezcla principal se puede aumentar de escala, p. ej., 10% extra para compensar las pérdidas durante el pipeteo, según el número de muestras.

15 Tabla 6 Componentes de la mezcla principal de tagmentación

Componente	Volumen (µl)
Tampón de ADN de tagmentación	11
Mezcla Amplicon Tagment Nextera	2.2
Super Q H ₂ O	3.3
Total	16.5

20 El tampón de ADN de tagmentation y la mezcla Amplicon Tagment Nextera están disponibles en Illumina, Inc (San Diego, CA; componentes nº 15027866 y 15031561). El tampón de ADN de tagmentación incluye Tris(hidroximetil)aminometano, MgCl₂ y dimetilformamida. La mezcla de Amplicon Tagment Nextera incluye la enzima transposoma. Luego se pueden añadir 15 µl de la mezcla principal de tagmentación a cada lisado celular, p. ej., generado a partir del Ejemplo 1, e incubar con el lisado celular a 55°C durante 5 min, y luego a 4°C para terminar la reacción. Luego, se puede añadir al tubo tampón de tagmentación neutralizador que incluye SDS (disponible en Illumina, Inc, San Diego, CA) y se incuba a temperatura ambiente durante 5 minutos.

Ejemplo 3 Amplificación por PCR de ciclo limitado

30 Los fragmentos de ADN tagmentados, p. ej., como se preparó en el Ejemplo 2, pueden amplificarse por un programa de PCR de ciclo limitado. Esta etapa de PCR también puede añadir otras secuencias en los dos extremos de los fragmentos de ácido nucleico marcados, p. ej., índice 1 (i7) e índice 2 (i5) (disponibles en Illumina, Inc, San Diego, CA) y secuencias requeridas para otros fines, p. ej., formación de grupos. Por ejemplo, los siguientes componentes en la Tabla 7 (disponibles en Illumina, Inc, San Diego, CA) se pueden añadir a la tagmentación neutralizada producida a partir del Ejemplo 3.

Tabla 7 Componentes para PCR de ciclo limitado

Componente	Volumen (μ l)
Mezcla principal de PCR	15
Cebador de índice 1 (cebador P5)	5
Cebador de índice 2 (cebador P7)	5

La mezcla principal de PCR en la Tabla 7 se puede preparar como en la Tabla 8 a continuación:

Tabla 8 Componentes de la mezcla principal de PCR

Componente	Concentración madre	Concentración de mezcla principal	Volumen (μ l)
Tampón KAPA HiFi Fidelity	5X	3.33X	999
Mezcla de dNTP	25 mM cada uno	1.00 mM cada uno	59.94
ADN polimerasa KAPA HiFi	1 U/ μ l	0.033 U/ μ l	49.95
Super Q H ₂ O			391.11
Total			1500

5

Un programa de PCR de ejemplo es el siguiente: 72°C 3 min, 98°C 30 segundos, y luego 20 ciclos de 98°C 10 segundos, 60°C 30 segundos y 72°C 30 segundos, y finalmente las muestras se mantienen a 4°C.

Ejemplo 4 La actividad de proteasa es útil para el acceso uniforme al ADN

10 El efecto de la actividad de la proteasa sobre el acceso uniforme al ADN se analiza en este ejemplo. En particular, se usan proteasas de 0 mg/ml, 0.1 mg/ml (4.5 mUA/ml), 0.5 mg/ml (22.5 mUA/ml) o 2.5 mg/ml (112.5 mUA/ml) para tratar células completas y núcleos. El porcentaje de lecturas mapeadas únicas se analiza para cada secuenciación. La Figura 1 es un histograma que muestra el porcentaje de lecturas mapeadas únicas en una secuenciación usando 0 mg/ml, 0.1 mg/ml, 0.5 mg/ml o 2.5 mg/ml de células enteras o núcleos tratados con proteasas. Como se muestra, el porcentaje de lecturas mapeadas únicas aumenta a medida que aumenta la concentración de proteasa, y esto es cierto utilizando tanto células enteras como núcleos como material de partida. También se observa que el porcentaje de lecturas mapeadas únicas que usa 0.5 mg/ml de proteasa es similar al que usa 2.5 mg/ml.

15 El efecto de la actividad de la proteasa sobre el acceso uniforme al ADN se analiza adicionalmente comparando los resultados de los análisis de recuentos y números de copias entre el uso del control de ADN genómico en masa con la preparación de la biblioteca Nextera XT, el uso de una sola célula con suficiente actividad de proteasa y el uso de una sola célula con insuficiente actividad de proteasa. La Figura 2 muestra histogramas de resultados de recuentos y análisis de número de copias usando ADN en masa, célula individual tratada con suficiente actividad de proteasa y célula individual tratada con actividad de proteasa insuficiente. Como se muestra, cuando se usa una cantidad relativamente grande de ADN genómico con el método actual de preparación de la biblioteca Nextera XT, como se muestra en el panel superior de la Figura 2, los resultados del análisis del número de copias limpias se pueden lograr con un ruido insignificante. Cuando solo se usa una sola célula para la secuenciación, el ruido es significativo y los datos del análisis del número de copias muestran un patrón de distribución disperso como se muestra en el panel inferior de la Figura 2. Sorprendentemente, cuando la célula individual se trata con suficiente proteasa (0.5 mg/ml), los resultados del análisis del número de copias se restablecen para ser comparables con los que usan ADN genómico en masa, mostrando datos limpios con ruido insignificante, como se muestra en el panel central de la Figura 2. Esto indica que la proteasa puede aumentar la accesibilidad del ADN genómico por transposasa ya que las proteínas de unión al ADN pueden eliminarse de manera uniforme.

Estos resultados muestran que la actividad de la proteasa es útil para el acceso uniforme al ADN en la secuenciación.

Ejemplo 5 Optimización de la concentración de proteasa

35 En este ejemplo, se analiza la concentración de proteasa utilizada en el presente método. La Figura 3A muestra histogramas de resultados de análisis de número de copias en una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 2 mg/ml o proteasa activa 2 mg/ml. Como se muestra, cuando la célula individual se trata con proteasa activa 0.5 mg/ml o 2 mg/ml, el resultado del análisis del número de copias limpias se logra de manera similar, como se muestra en los dos histogramas superiores de la Figura 3A. En contraste, cuando la reacción se realiza con proteasa precalentada inactivada a 70°C, no se puede lograr un resultado de número de copias limpio, como se muestra en el histograma inferior de la Figura 3A. Este resultado muestra que la proteasa de las concentraciones tanto 0.5 mg/ml como 2 mg/ml es efectiva y suficiente.

40

El porcentaje de lecturas mapeadas únicas también se analiza en una secuenciación de una sola célula tratada con

proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml, proteasa activa 2 mg/ml o proteasa previamente inactivada por calor (a 70°C) 2 mg/ml. La Figura 3B muestra un histograma de porcentaje de lecturas mapeadas únicas en una secuenciación de una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml, proteasa activa 2 mg/ml o proteasa previamente inactivada por calor 2 mg/ml y una muestra de control sin células. Como se muestra, los porcentajes de lecturas mapeadas únicas en la secuenciación usando una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml y proteasa activa 2 mg/ml son aproximadamente de 65% con una pequeña variación. En contraste, cuando la proteasa se inactiva por debajo de 70°C, incluso si se usa una mayor cantidad de proteasa, el porcentaje de lecturas mapeadas únicas es mucho menor con grandes variaciones.

Además, el ruido en los datos del número de copias se analiza analizando las diferencias de recuento entre el recuento de clases adyacentes. La Figura 3C muestra un histograma de las diferencias de recuentos de lecturas entre clases adyacentes (intervalo intercuartílico de diferencia de recuentos de lecturas entre clases adyacentes) en una secuenciación de una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml, proteasa activa 2 mg/ml o proteasa previamente inactivada por calor 2 mg/ml, y una muestra de control sin células. Como se muestra, las diferencias de recuento entre el recuento de clases adyacentes en una secuenciación usando una sola célula tratada con proteasa activa 0.5 mg/ml, proteasa activa 1 mg/ml y proteasa activa 2 mg/ml son relativamente pequeñas (aproximadamente 20%) con pequeñas variaciones. Por el contrario, cuando la proteasa se inactiva por debajo de 70°C, incluso si se usa una mayor cantidad de proteasa (2 mg/ml), la diferencia de recuentos entre recuentos de clases adyacentes es mucho mayor con grandes variaciones.

En conjunto, estos resultados muestran que la proteasa con un intervalo de concentración de 0.5 mg/ml a 2.0 mg/ml (22.5 mUA/ml a 90 mUA/ml) es suficiente y efectiva en el método proporcionado en el presente documento.

Ejemplo 6 Optimización de las condiciones de pH de la reacción de digestión de proteasa

En este ejemplo, las condiciones de pH de la reacción de digestión de la proteasa se optimizan equilibrando la actividad de la proteasa y los resultados de secuenciación.

La actividad de la proteasa se analiza en diferentes condiciones de pH. El resultado se muestra en la Figura 4A. La Figura 4A es un histograma que muestra la actividad relativa (relativa a la actividad de la proteasa a pH 8.0) de la proteasa a pH 7.0, pH 7.5, pH 8.0, pH 8.5, pH 9.0 o pH 10.0. Como se muestra, la actividad de la proteasa aumenta a medida que el valor del pH aumenta teniendo la proteasa la actividad más baja a pH 7.0 y la actividad más alta a pH 10.0.

Después se analiza el porcentaje de lecturas mapeadas únicas en varias condiciones de pH. La Figura 4B muestra un histograma del porcentaje de lecturas mapeadas únicas en una secuenciación de una sola célula tratada con proteasa a pH 7.0, pH 8.0, pH 9.0 o pH 10.0. Como se muestra, cuando el pH es 7, 8 o 9, se puede lograr aproximadamente 70% de lecturas mapeadas únicas y limpias. Sin embargo, cuando el pH es 10, se puede lograr menos porcentaje de lecturas mapeadas únicas y la variación de datos aumenta significativamente.

El ruido en los datos del número de copias también se analiza comparando las diferencias de recuentos entre clases adyacentes. La Figura 4C muestra un histograma de las diferencias de recuentos de lecturas entre clases adyacentes (intervalo intercuartílico de diferencia de recuentos de lecturas entre clases adyacentes) en una secuenciación de una sola célula tratada con proteasa de 0.5 mg/ml a pH 7.0, pH 8.0, pH 9.0 o pH 10.0. Como se muestra, de acuerdo con los resultados de lecturas mapeadas únicas, las diferencias de recuentos entre clases adyacentes son relativamente pequeñas (aproximadamente 20%) con pequeñas variaciones; mientras que las diferencias de recuentos entre clases adyacentes aumentan significativamente con una gran variación a pH 10.0.

En algunas realizaciones, el valor de pH de la reacción de digestión está entre pH 7.0 y pH 9.0.

Ejemplo 7 Ensayo de inactivación por calor de proteasa

En algunas realizaciones, la proteasa proporcionada en el presente documento puede inactivarse por calor. Como se ha descrito anteriormente, en las realizaciones preparadas, la proteasa se puede inactivar a una temperatura relativamente baja (p. ej., 70°C) para que la conformación de ADN bicatenario se pueda preservar para la reacción de tagmentación. En este ejemplo, se analiza la inactivación por calor de la proteasa (de QIAGEN) y su efecto sobre los resultados de secuenciación.

La proteasa se calentó previamente a diferentes temperaturas y se ensayó la actividad de la proteasa. El resultado se muestra en Figura 5A, que muestra un histograma de actividad de proteasa relativa cuando se precalienta a temperatura ambiente, 50°C, 60°C o 70°C. Como se muestra, la actividad de la proteasa disminuye progresivamente a medida que aumenta la temperatura y se inactiva por completo a 70°C. Este resultado es consistente con los resultados mostrados en el Ejemplo 5 anterior.

Se analiza el porcentaje de lecturas mapeadas únicas en la secuenciación de una sola célula, tres células y 15 pg de ADN genómico a varias temperaturas. La Figura 5B muestra un histograma de porcentaje de lecturas mapeadas únicas en una secuenciación de una sola célula, tres células o 15 pg de ADN genómico, tratado con proteasa 2.0 mg/ml a temperatura ambiente, 50°C, 60°C o 70°C. Como se muestra, el porcentaje de lecturas mapeadas únicas disminuye a medida que aumenta la temperatura. Sin embargo, debido a que se usa una concentración relativamente

mayor de proteasa (2.0 mg/ml) en el experimento, hay más tolerancia para la actividad reducida de proteasa a 70°C. Como tal, el porcentaje de lecturas mapeadas únicas a 70°C sigue siendo relativamente alto, aunque más bajo que los tratados a temperaturas más bajas.

5 También se analizan las diferencias de recuento entre clases adyacentes en la secuenciación de una sola célula, tres células y 15 pg de ADN genómico a varias temperaturas. La Figura 5C muestra un histograma de diferencias de recuentos de lecturas entre clases adyacentes (intervalo intercuartílico de diferencia de recuentos de lecturas entre clases adyacentes) en una secuenciación de una sola célula, tres células o 15 pg de ADN genómico, tratados con proteasa 2 mg/ml a temperatura ambiente, 50°C, 60°C o 70°C. Como se muestra, las diferencias de recuentos entre clases adyacentes son relativamente pequeñas con pequeñas variaciones a temperaturas más bajas (p. ej., a 10 temperatura ambiente y 50-60°C); mientras que las diferencias de recuentos entre clases adyacentes aumentan significativamente con una mayor variación a 70°C.

Ejemplo 8 La diversidad de la biblioteca aumenta con tamaños de insertos más pequeños

En una secuenciación de una sola célula, solo están presentes dos copias del genoma y, por lo tanto, un tamaño de inserto más pequeño tiende a aumentar la diversidad de la biblioteca. Como se muestra en la Figura 6A, los recuentos, 15 y por lo tanto la diversidad representada por una biblioteca, aumentan a medida que disminuye el tamaño del inserto. Por lo tanto, en algunas realizaciones, el método en el presente documento usa una mayor cantidad de transposasa en la etapa de tagmentación para aumentar la fragmentación y reducir el tamaño del inserto de los fragmentos de ácido nucleico marcados. La Figura 6B muestra el tamaño del inserto de una biblioteca tratada con 1 µl de Tn5 o 2 µl de Tn5. Como se muestra, cuando se usa 1 µl de Tn5 en una reacción de tagmentación, el tamaño medio del fragmento es de aproximadamente 550 pb; mientras que cuando se usan 2 µl de Tn5 en una reacción de tagmentación, el tamaño 20 medio del fragmento es de aproximadamente 400 pb. De acuerdo con un tamaño de inserto más pequeño, la diversidad de la biblioteca aumenta cuando se trata con 2 µl de Tn5 en comparación con el tratado con 1 µl de Tn5, como se muestra en la Figura 6C.

Ejemplo 9 Optimización de los ciclos de PCR

En una secuenciación que utiliza una población mínima de células, el ADN de entrada es relativamente pequeño y, por lo tanto, el número de ciclos de PCR se puede ajustar para lograr mejores resultados de secuenciación. En este ejemplo, el número de ciclos de PCR se ensaya y optimiza utilizando una sola célula como material de partida. La Figura 7 muestra histogramas de recuentos y resultados de análisis de número de copias en una secuenciación de una sola célula de acuerdo con el método proporcionado en el presente documento usando PCR con 16 ciclos, 18 30 ciclos o 20 ciclos. Como se muestra, el ruido es grande cuando se usa PCR con 16 ciclos, y el ruido se reduce significativamente cuando se usa PCR con 18 ciclos o 20 ciclos.

Ejemplo 10 Distribución de lecturas usando una, tres o cinco células

La distribución de lecturas usando una, tres o cinco células se analiza en este ejemplo. La Figura 8A muestra la 35 distribución de lecturas de tres secuenciaciones de una sola célula. Como se muestra, las regiones de lectura no se solapan completamente entre las tres secuenciaciones de una sola célula. Por lo tanto, aumentar el número de células puede ayudar a una cobertura más amplia. La Figura 8B muestra la distribución de lecturas de la secuenciación de una sola célula, secuenciación de tres células o secuenciación de cinco células. Como se muestra, la cobertura genómica aumenta a medida que aumenta el número de células. La Figura 8C muestra histogramas de diversidad de biblioteca promedio y cobertura genómica estimada usando una sola célula, tres células o cinco células. Como se muestra, se estima que una célula puede cubrir aproximadamente 40% del genoma, y tres células pueden cubrir más 40 de 50% del genoma, y cinco células pueden cubrir aproximadamente 60% del genoma. Los recuentos medios de la biblioteca con una célula, tres células y cinco células es de aproximadamente 5 millones, 15 millones y 20 millones, respectivamente.

La figura 8D muestra la tasa de éxito general. Como se muestra, cuando se usa más de una célula, la tasa de éxito general es de 99% (N = 81). Cuando se usa una sola célula, la tasa de éxito general también es relativamente alta 45 94% (N = 187).

Ejemplo 11 Comparación de recuentos y datos de números de copias entre diferentes métodos de preparación de bibliotecas

En este ejemplo, el método proporcionado en el presente documento se compara con algunos métodos actuales de 50 preparación de células individuales.

La Figura 9A muestra el análisis del número de copias usando REPLI-g de célula única (MDA) con la preparación de la biblioteca Nextera XT. El kit REPLI-g de célula única desarrollado por QIAGEN está especialmente diseñado para amplificar el ADN genómico de células individuales (1 a <1000 células) o ADN genómico purificado con cobertura genómica. El kit REPLI-g de célula única desarrollado por QIAGEN utiliza la tecnología de amplificación por desplazamiento múltiple (MDA). Véase Spits et al., 2006, Whole-genome multiple displacement amplification from single cells, *Nature protocols* 1 (4): 1965-70. Sin embargo, debido a que la MDA introducía un sesgo de 55 sobreamplificación, los datos de variación del número de copias tienen mucho ruido cuando derivan de una sola célula,

tres células o cinco células, como se muestra en la Figura 9A.

5 La Figura 9B muestra el análisis del número de copias usando SurePlex (PicoPlex) con la preparación de la biblioteca Nextera XT. El sistema de amplificación SurePlex desarrollado por Illumina, Inc (San Diego, CA) es una solución para la extracción y amplificación de ADN de una o pocas células. Como se muestra, el sistema de amplificación SurePlex reduce significativamente el ruido en comparación con MDA.

La Figura 9C muestra el análisis del número de copias usando un método (Nextera SC) proporcionado en el presente documento. Como se muestra, el ruido se reduce aún más en comparación con el uso del sistema de amplificación SurePlex.

Ejemplo 12 Detección de mosaicismo

10 En este ejemplo, se ilustra el uso del método proporcionado en el presente documento para detectar mosaicismo. La Figura 10A muestra datos de análisis del número de copias del cromosoma 18 usando una sola célula GM50121. Se muestran los datos del número de copias de tres secuenciaciones de una sola célula. Se detecta una población que representa ADN de 15.4 MB en cada secuenciación de una sola célula. La Figura 10B muestra los datos del número de recuentos del uso de una sola célula GM20916. Como se muestra, las flechas indican los recuentos originados por el mosaicismo. La Figura 10C muestra datos de análisis de número de copias de los cromosomas 15, X y 10 usando una sola célula GM20916. Los datos del número de copias para cada cromosoma analizado detectan una población adicional que representa otro cromosoma. De manera similar, la Figura 10D muestra datos de análisis de número de copias de los cromosomas 1 y 11 usando una sola célula GM10239. Como se muestra en estas figuras, los datos del número de copias para cada cromosoma analizado en la Figura 10D también detecta una población adicional que representa otro cromosoma.

15

20

REIVINDICACIONES

1. Un método para preparar una biblioteca de fragmentos de ácido nucleico marcados que comprende:
 - (a) poner en contacto una célula individual directamente con un reactivo de lisis para generar un lisado celular, en donde el reactivo de lisis tiene una o más proteasas, y en donde el lisado celular contiene un ácido nucleico objetivo;
 - 5 (b) inactivar la una o más proteasas para formar un lisado celular inactivado, y
 - (c) aplicar directamente al menos una transposasa y al menos una composición de extremos de transposón que contiene una cadena transferida al lisado celular inactivado en condiciones donde el ácido nucleico objetivo y la composición de extremos de transposón experimentan una reacción de transposición para generar una mezcla, en donde:
 - 10 el ácido nucleico objetivo se fragmenta para generar una pluralidad de fragmentos de ácido nucleico objetivo, la cadena transferida de la composición de extremos de transposón se une a los extremos 5' de cada uno de una pluralidad de fragmentos de ácido nucleico objetivo para generar una pluralidad de fragmentos de ácido nucleico objetivo marcados en 5';
 - 15 el ácido nucleico objetivo es un ADN bicatenario, y en donde el ácido nucleico objetivo permanece como el ADN bicatenario antes de aplicar una transposasa y una composición de extremos de transposón en la etapa (c).
2. El método de la reivindicación 1, en donde las etapas (a), (b) y (c) se realizan en un solo tubo de reacción.
3. El método de la reivindicación 1, en donde el reactivo de lisis comprende uno o más detergentes.
4. El método de la reivindicación 1, en donde no se produce una etapa previa de purificación o amplificación de ADN entre (a) y (c).
5. El método de la reivindicación 4, en donde el ácido nucleico objetivo es ADN genómico, contiene ADN cromosómico o un fragmento del mismo, o comprende un genoma o un genoma parcial.
6. El método de la reivindicación 1, en donde la al menos una transposasa es una transposasa Tn5, la al menos una composición de extremos de transposón comprende un extremo del transposón Tn5, o
 - 25 en donde la cadena transferida comprende dominios de marcador que contienen uno o más de un dominio de sitio de restricción, un dominio de marcador de captura, un dominio de marcador de secuenciación, un dominio de marcador de amplificación, un dominio de marcador de detección y un dominio de marcador de dirección.
7. El método de la reivindicación 1, que comprende, además:
 - 30 (d) incubar la mezcla de la etapa (c) directamente con al menos una enzima modificadora de ácido nucleico en condiciones en las que un marcador de 3' se une a los fragmentos de ácido nucleico objetivo marcados en 5' para generar una pluralidad de fragmentos de ácido nucleico objetivo con dos marcadores, opcionalmente en donde las etapas (a), (b), (c) y (d) se realizan en un solo tubo de reacción.
8. El método de la reivindicación 7, en donde la enzima modificadora de ácido nucleico es una polimerasa y en donde dicho marcador de 3' se forma por extensión del extremo 3' del fragmento de ácido nucleico marcado en 5'.
9. El método de la reivindicación 8, en donde la enzima modificadora de ácido nucleico es una ligasa y en donde el marcador de 3' se forma por ligadura de un oligonucleótido al extremo 3' del fragmento de ácido nucleico marcado en 5', o que comprende además (e) amplificar uno o más fragmentos de ácido nucleico objetivo con dos marcadores para generar una biblioteca de fragmentos de ácido nucleico marcados con secuencia adicional en el extremo 5' y/o en el extremo 3' de los fragmentos de ácido nucleico con dos marcadores.
10. El método de la reivindicación 8, en donde las etapas (a), (b), (c), (d) y (e) se realizan en un solo tubo de reacción, y
 - (i) en donde la amplificación comprende el uso de una o más de una reacción en cadena de la polimerasa (PCR), una reacción de amplificación por desplazamiento de cadena, una reacción de amplificación por círculo rodante, una reacción en cadena de ligasa, una reacción de amplificación mediada por transcripción, o una reacción de amplificación mediada por bucle.
11. El método de la reivindicación 8 en donde
 - (i) la amplificación comprende una PCR que utiliza un solo cebador que es complementario del marcador de 3' de los fragmentos de ADN objetivo con dos marcadores, o

(ii) en donde la amplificación comprende una PCR usando un primer y un segundo cebador, en donde al menos una porción del extremo 3' del primer cebador es complementaria de al menos una porción del marcador de 3' de los fragmentos de ácido nucleico objetivo con dos marcadores, y en donde al menos una porción del extremo 3' del segundo cebador presenta la secuencia de al menos una porción del marcador de 5' de los fragmentos de ácido nucleico objetivo con dos marcadores.

5

12. El método de la reivindicación 11, en donde una porción del extremo 5' del primer cebador no es complementaria del marcador de 3' de los fragmentos de ácido nucleico objetivo con dos marcadores, y en donde una porción del extremo 5' del segundo cebador no presenta la secuencia de al menos una porción del marcador de 5' de los fragmentos de ácido nucleico objetivo con dos marcadores, opcionalmente en donde el primer cebador comprende una primera secuencia universal, y/o en donde el segundo cebador comprende una segunda secuencia universal.

10

13. El método de la reivindicación 1, que comprende además secuenciar los fragmentos de ácido nucleico marcados, y

(i) en donde la secuenciación de los fragmentos de ácido nucleico marcados comprende el uso de una o más de secuenciación por síntesis, PCR puente, secuenciación por terminación de cadena, secuenciación por hibridación, secuenciación por nanoporos y secuenciación por ligadura.

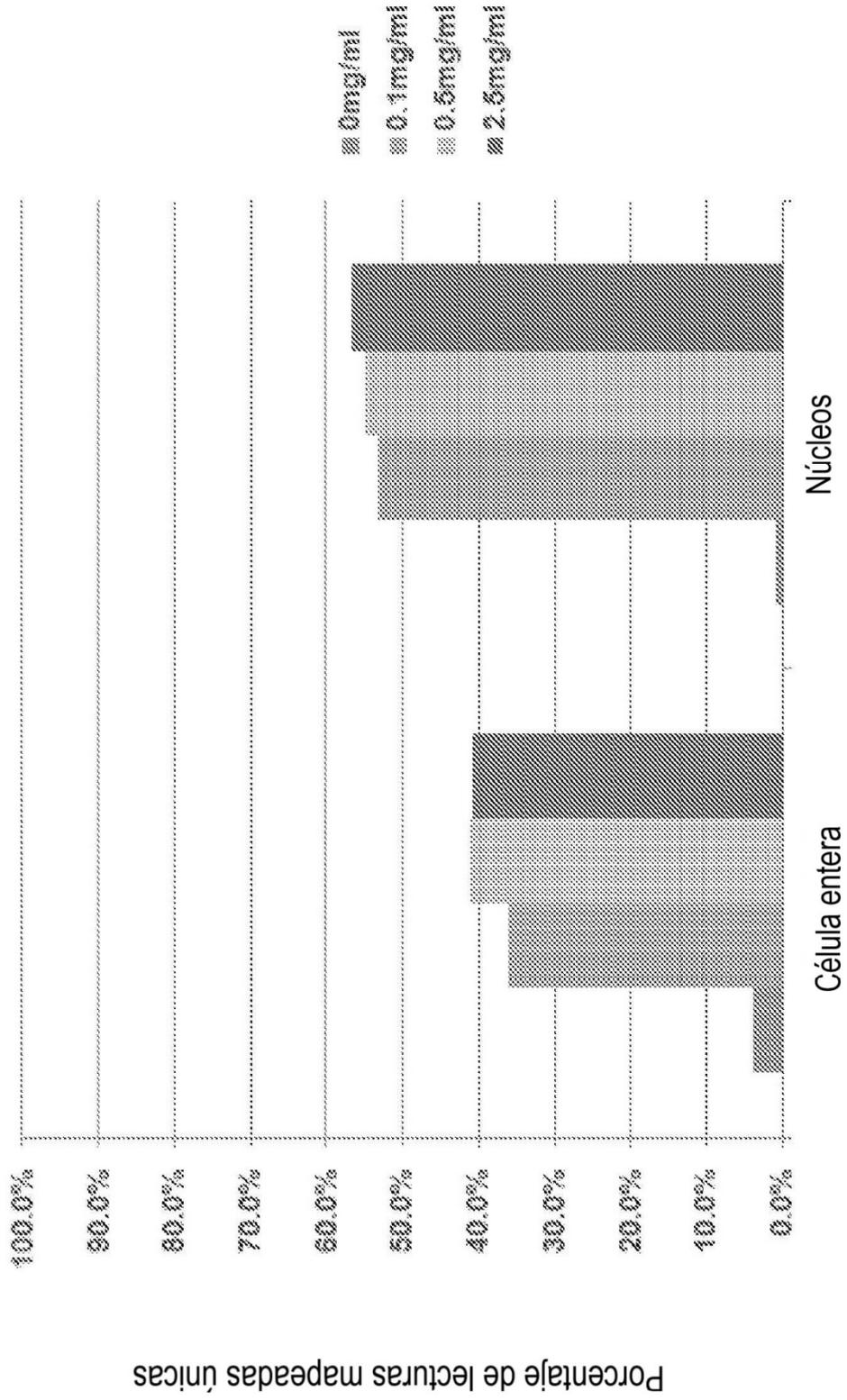


Figura 1



Figura 2

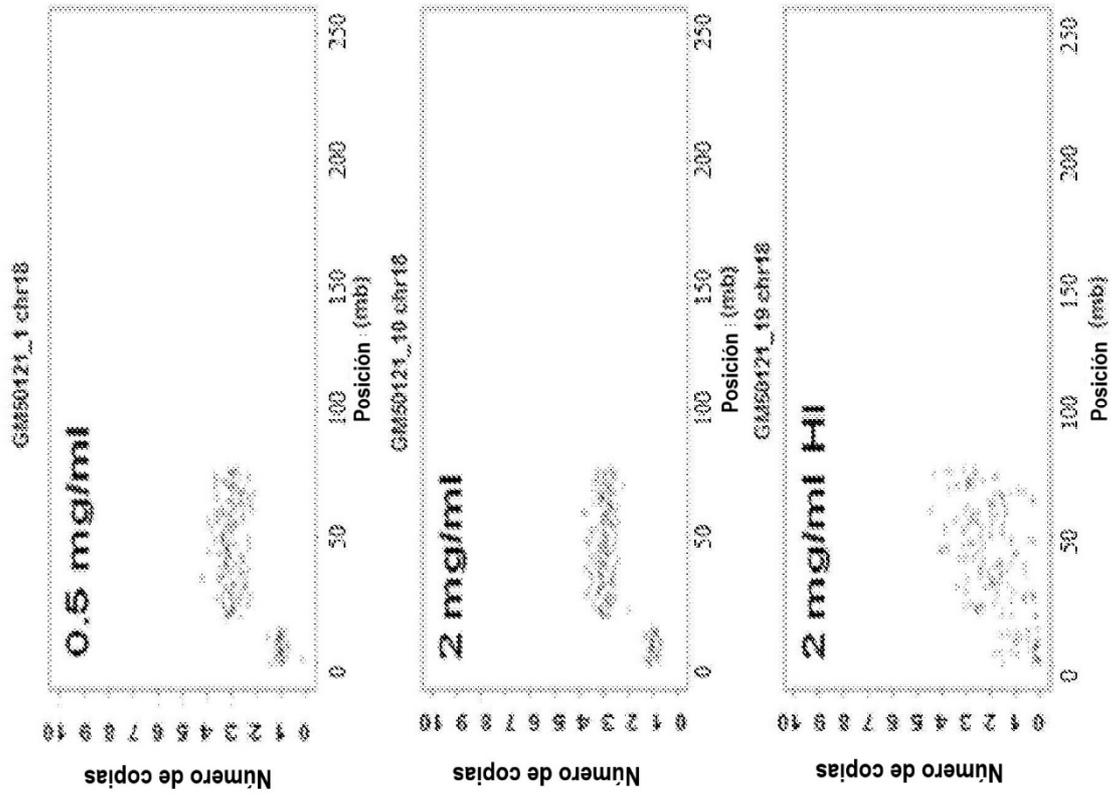


Figura 3A

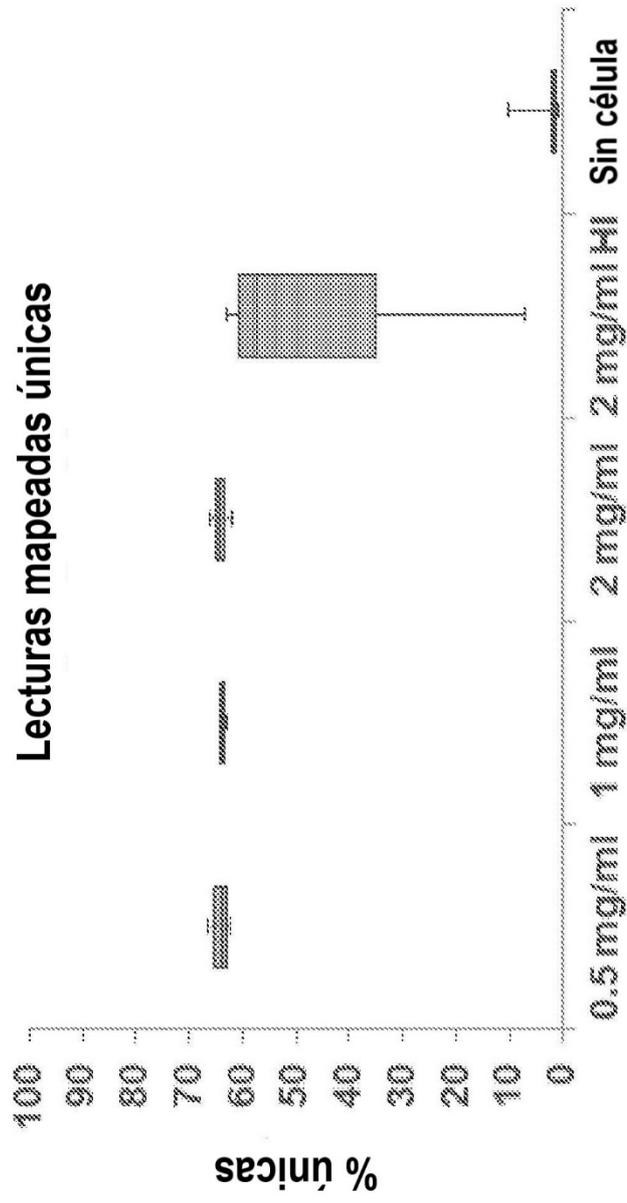


Figura 3B

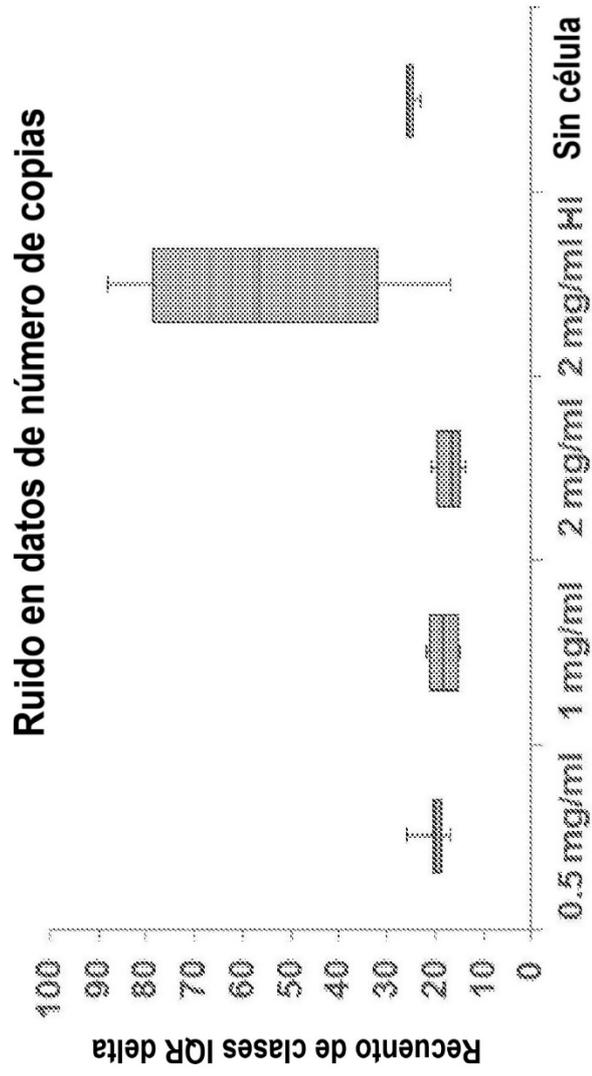


Figura 3C

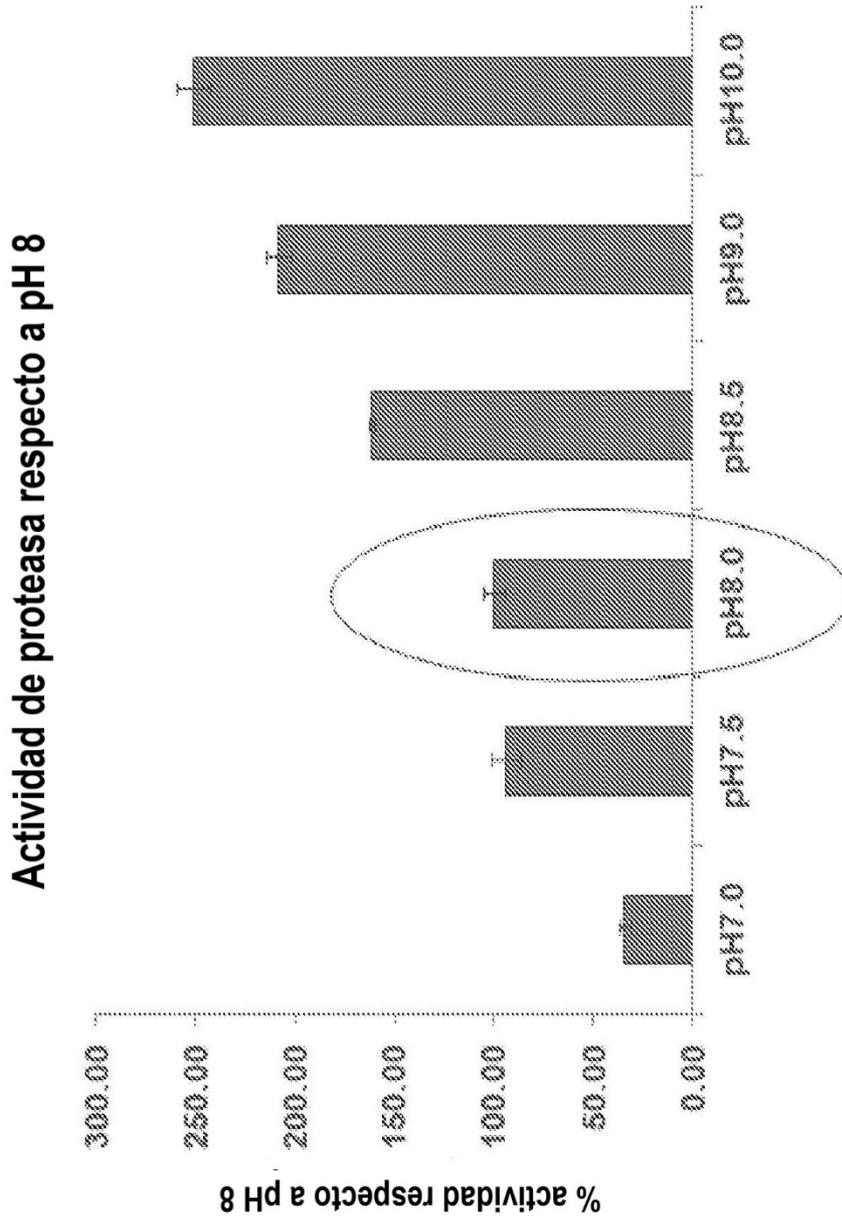


Figura 4A

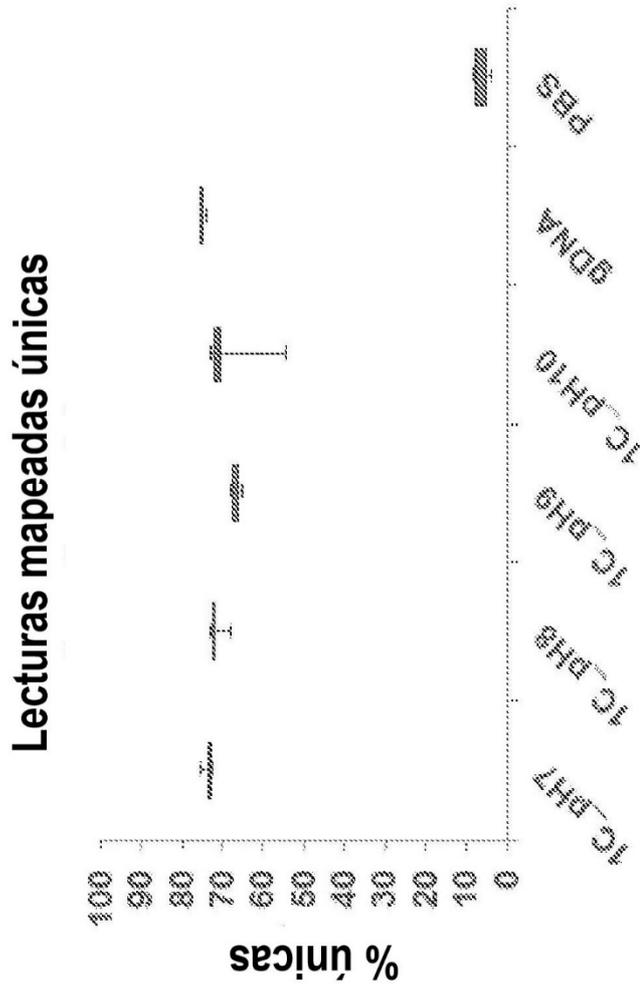


Figura 4B

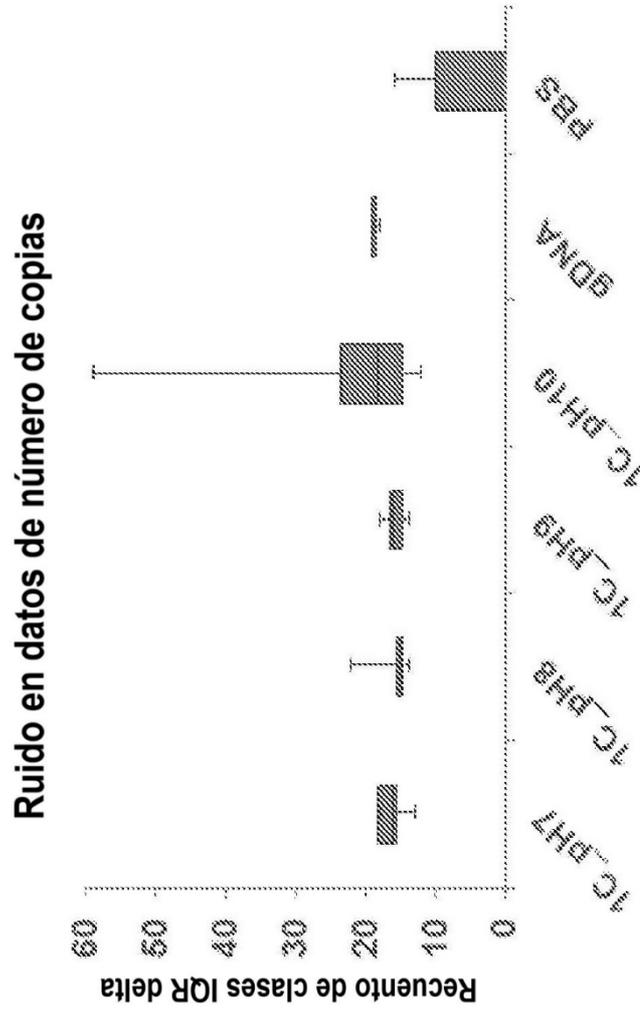


Figura 4C

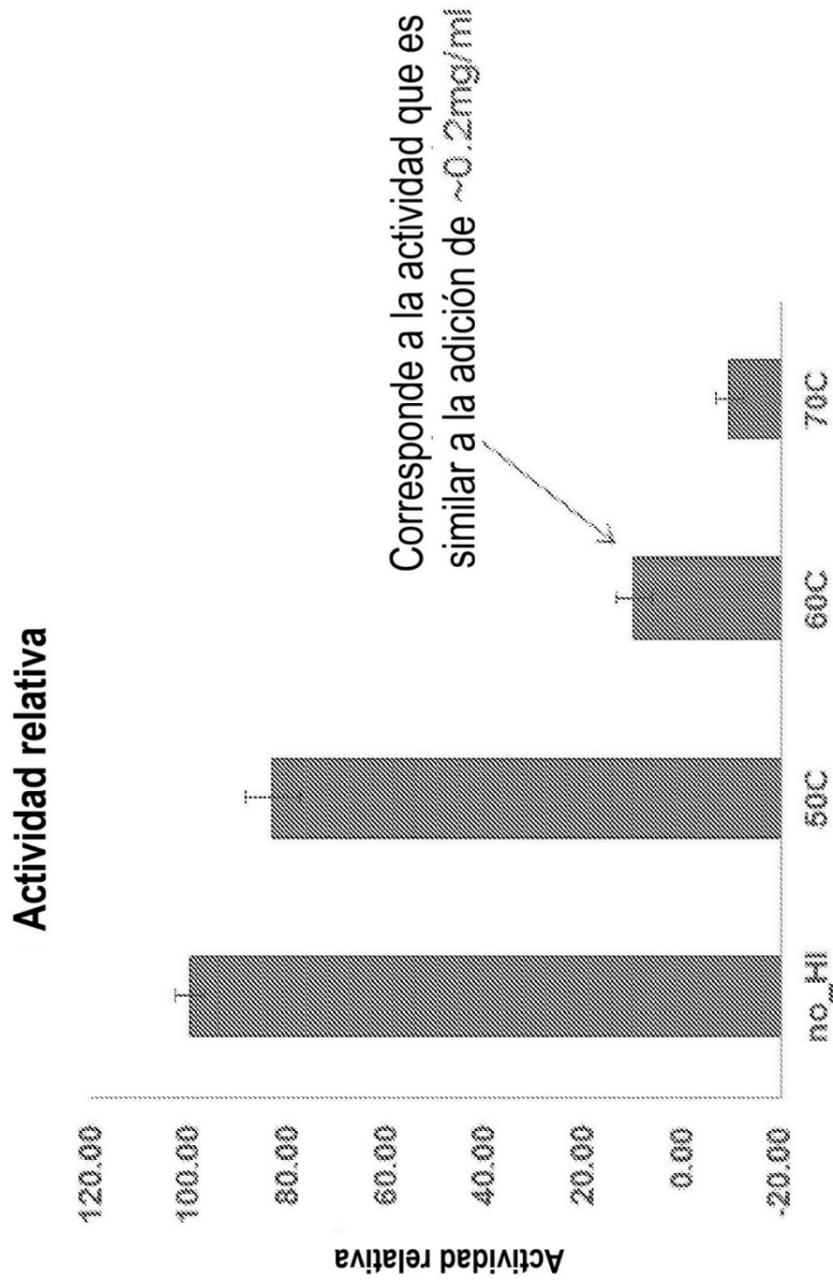


Figura 5A

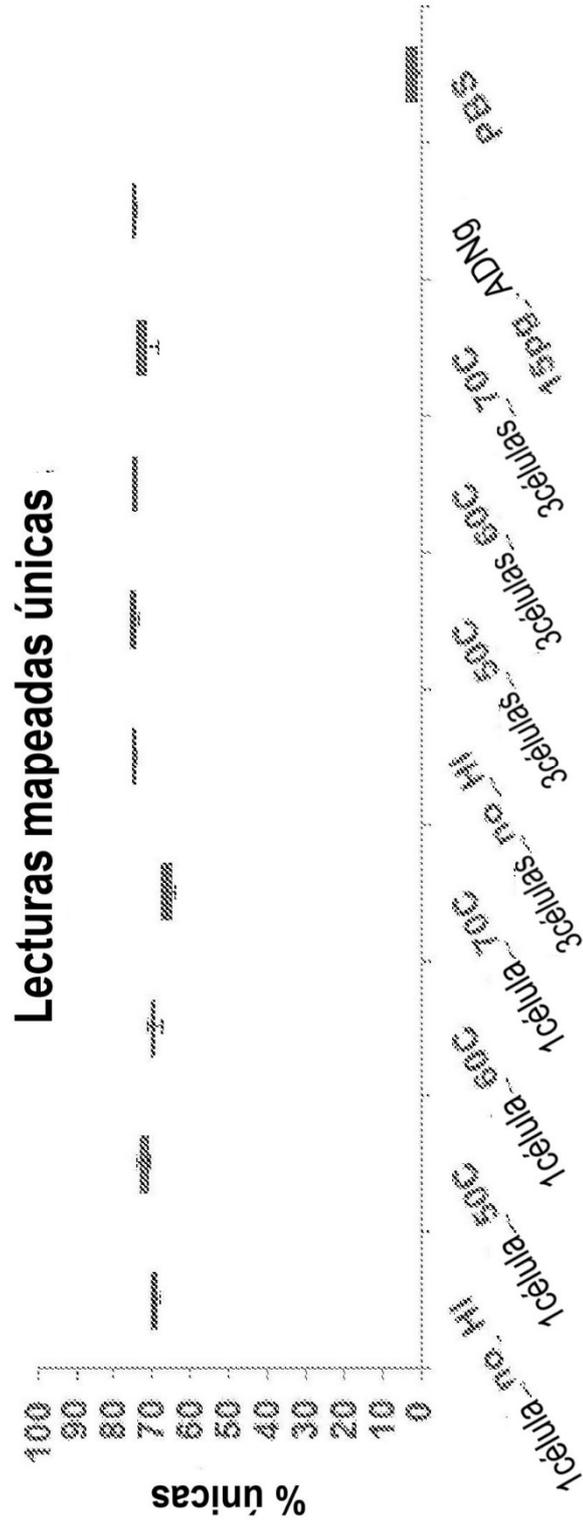


Figura 5B

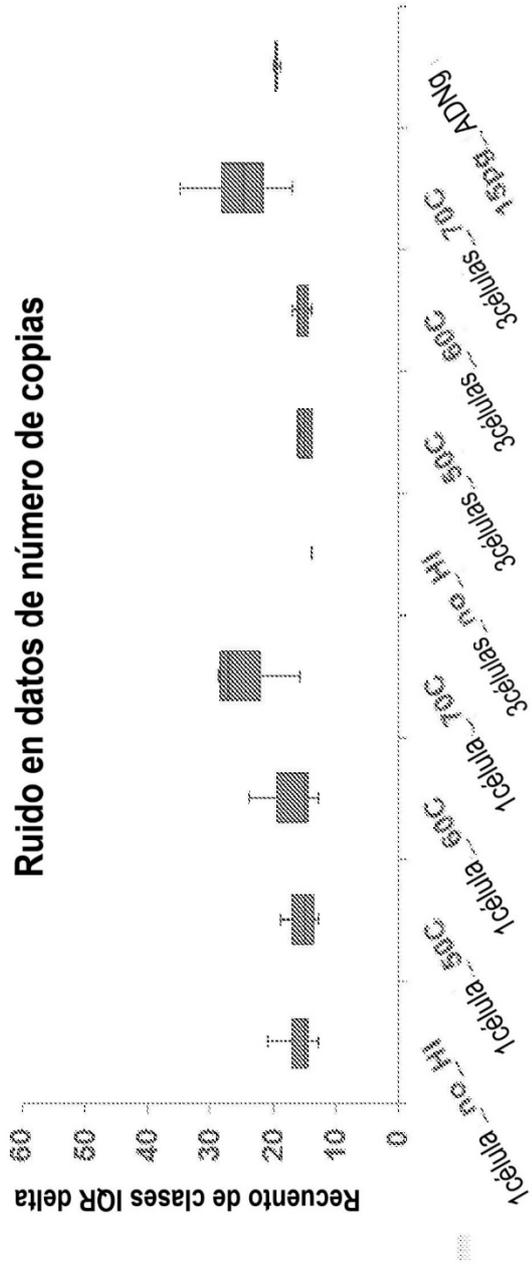


Figura 5C

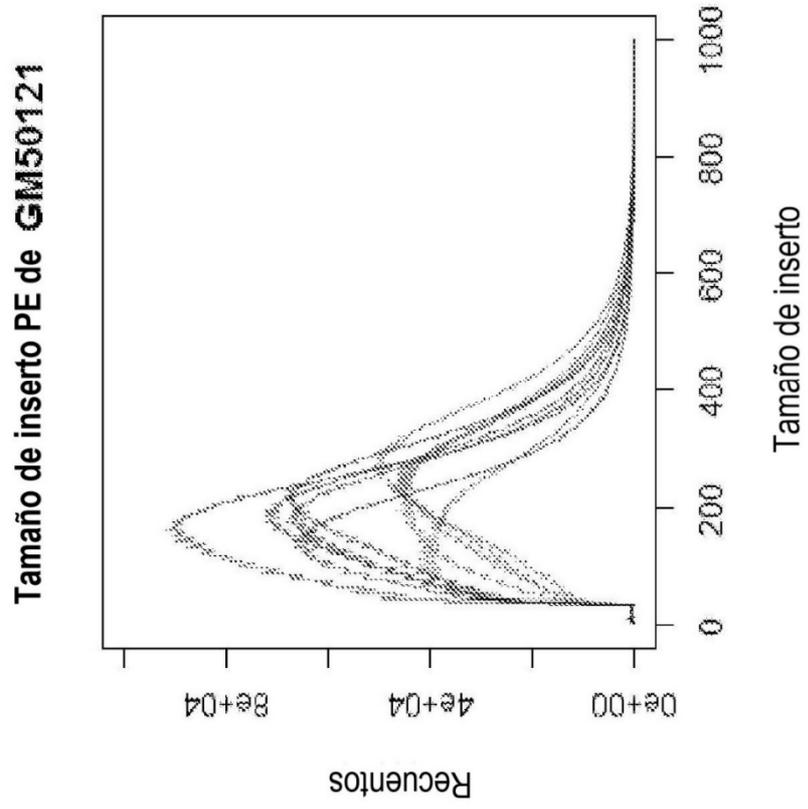


Figura 6A

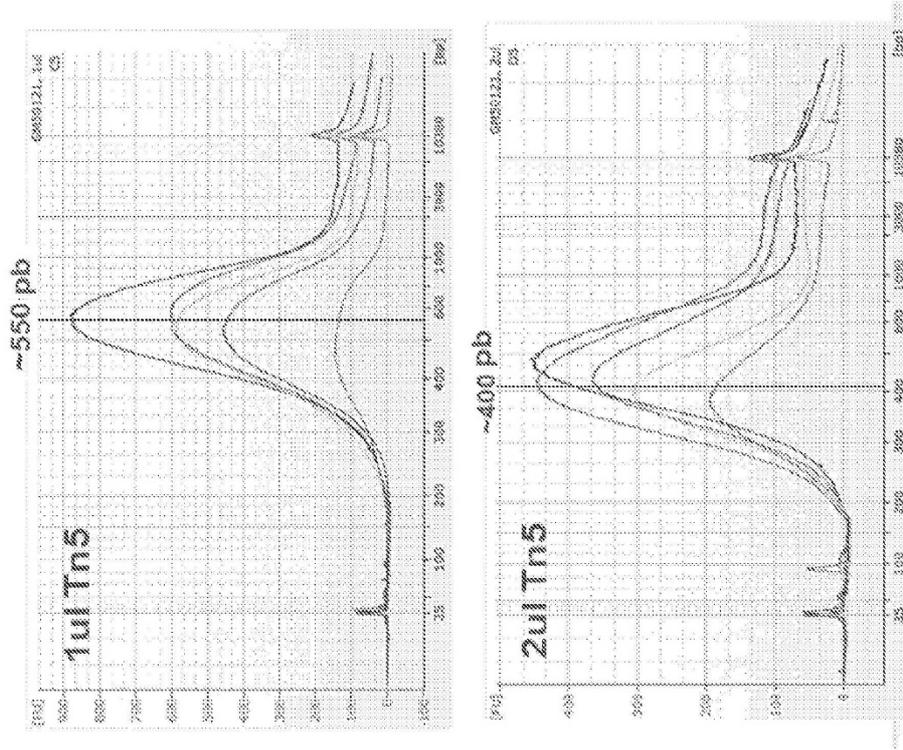
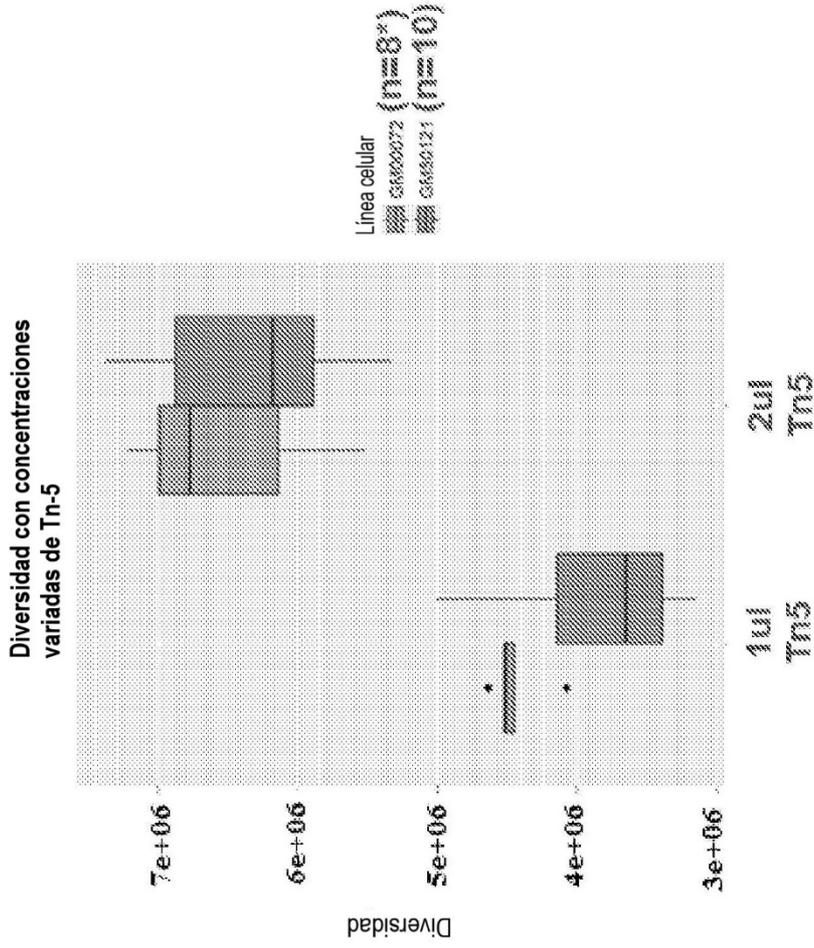


Figure 6B



- 1ul Tn5: diversidad < 5 millones
- 2ul Tn5: diversidad 6-7 millones

*2 muestras fallaron en la prep. de biblioteca

Figura 6C

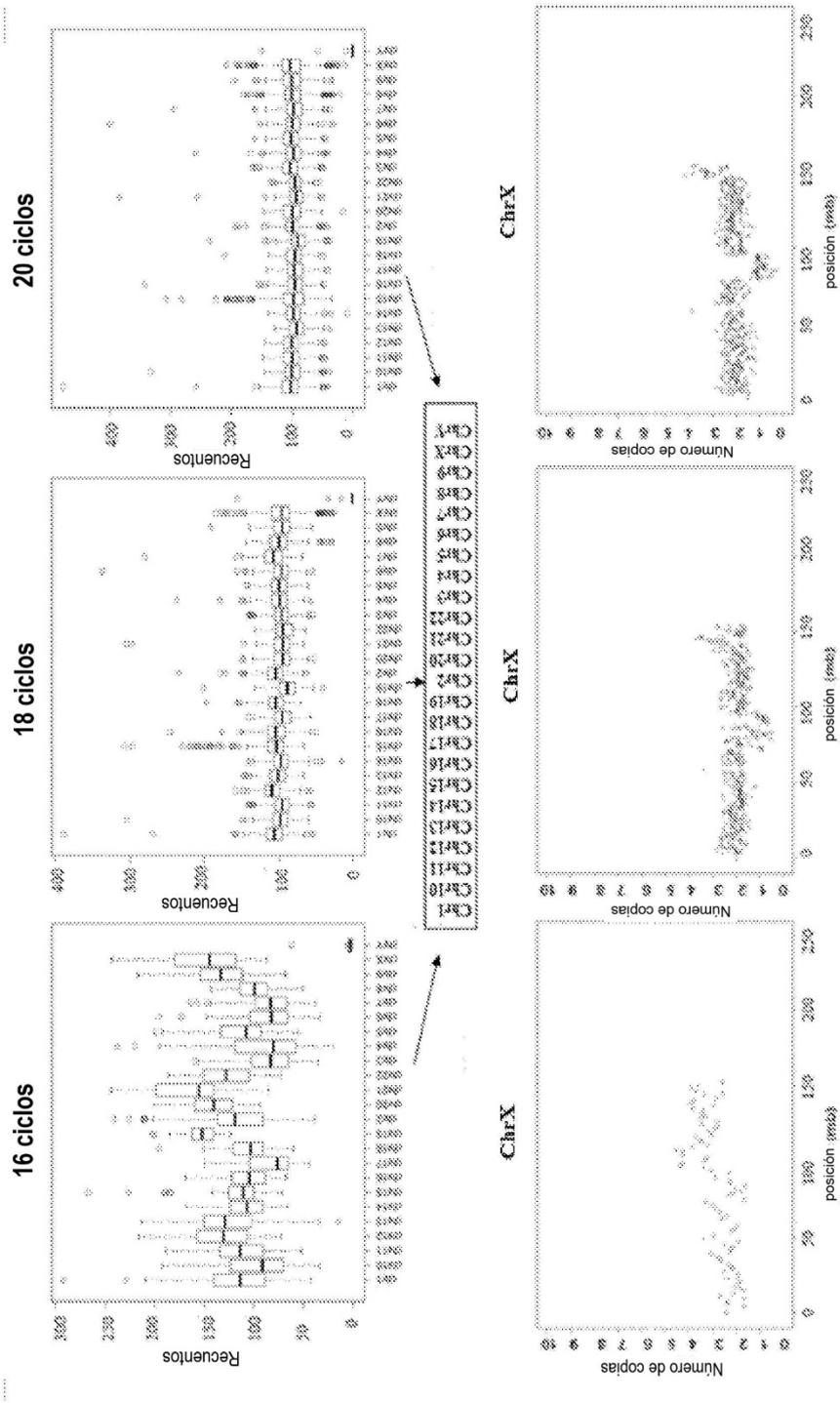


Figura 7

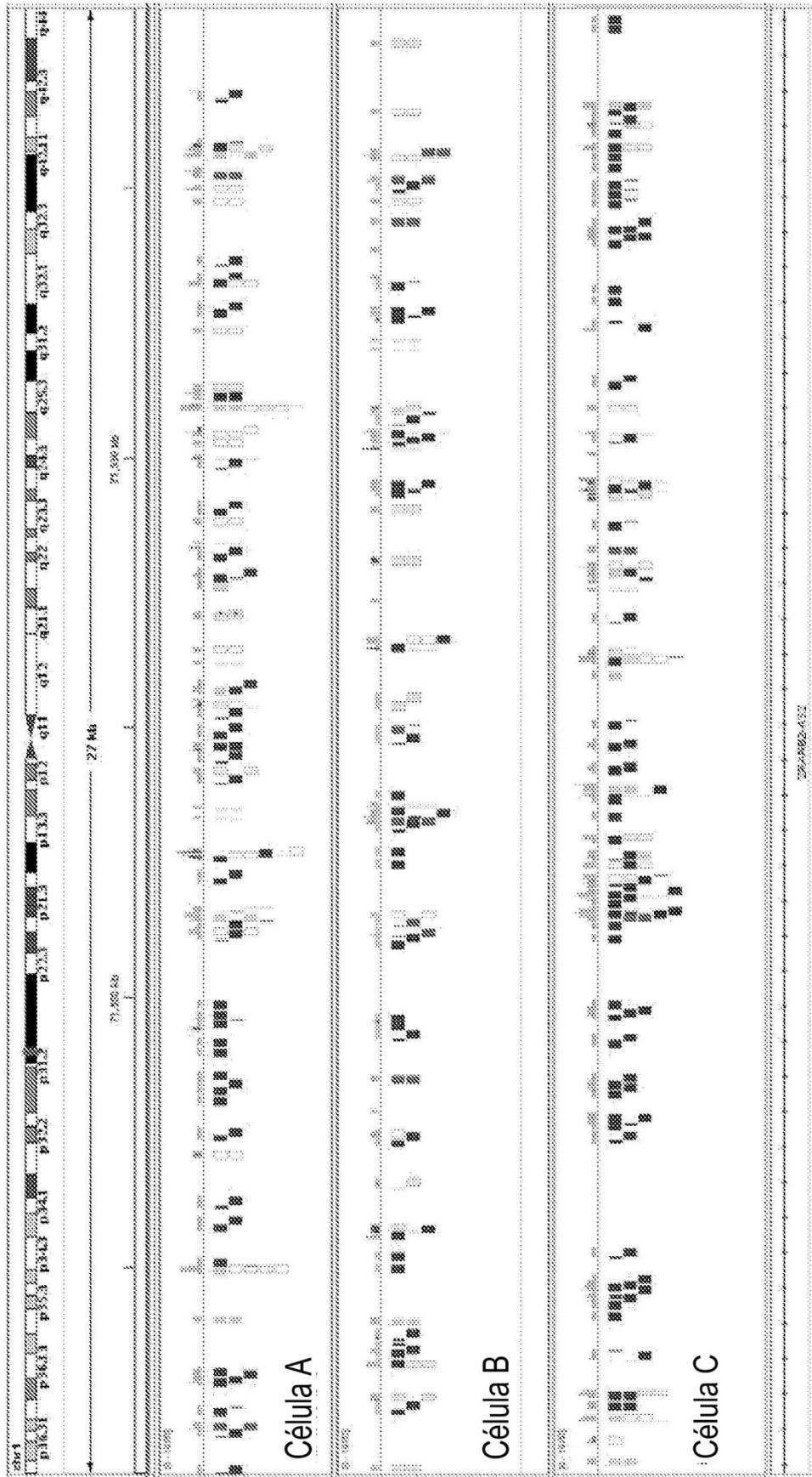


Figura 8A

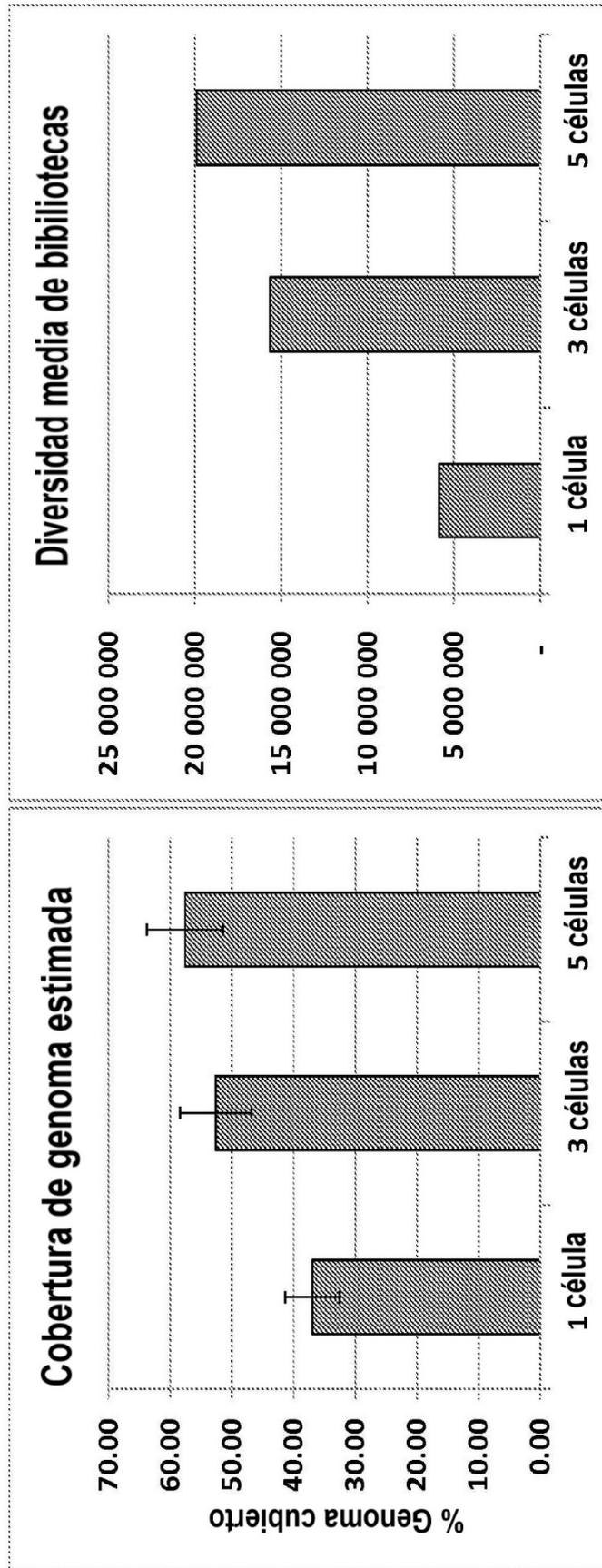


Figura 8C

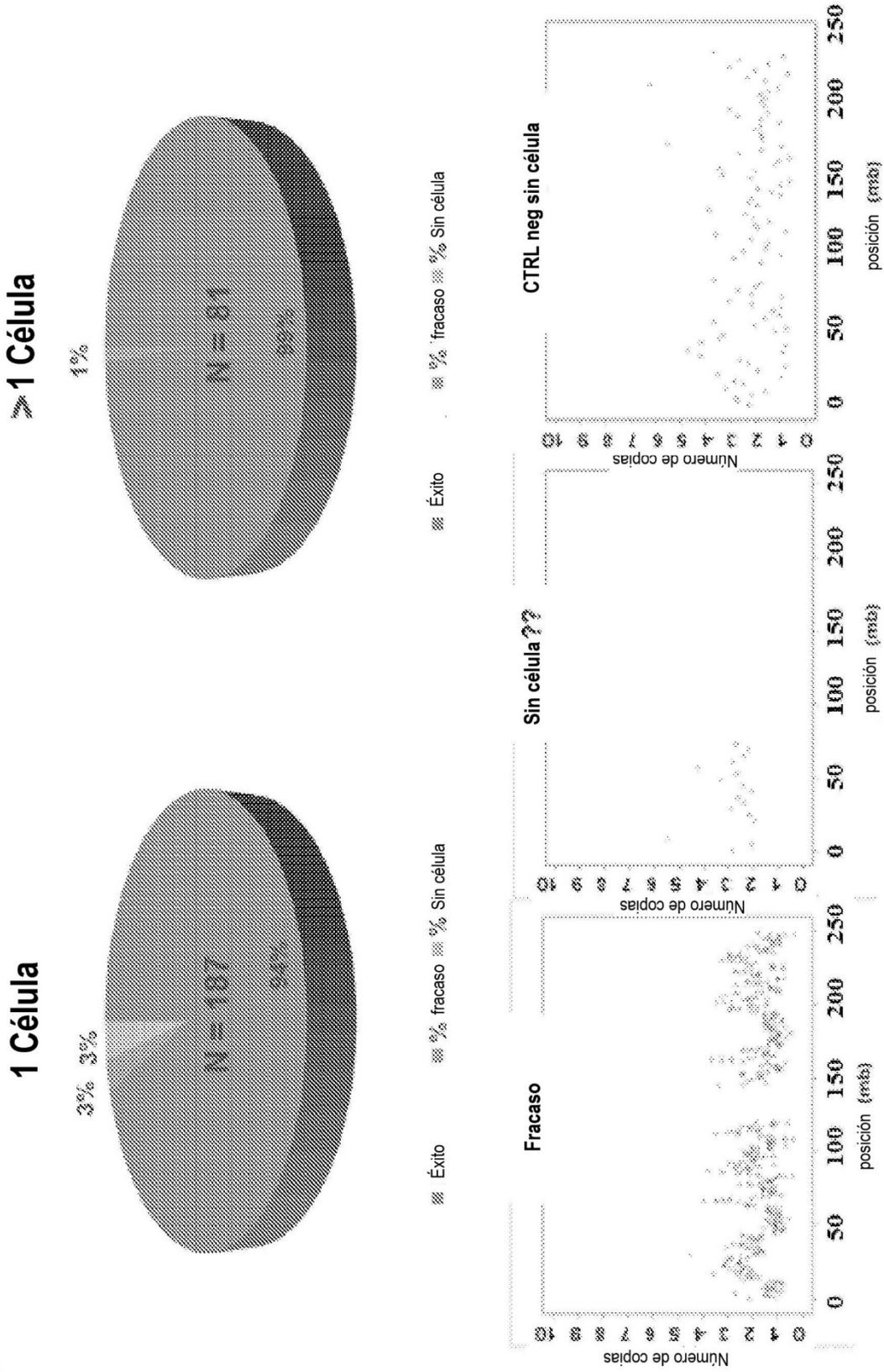


Figura 8D

REPLig de célula única (MDA) + Preparación de bibliotecas Nextera XT

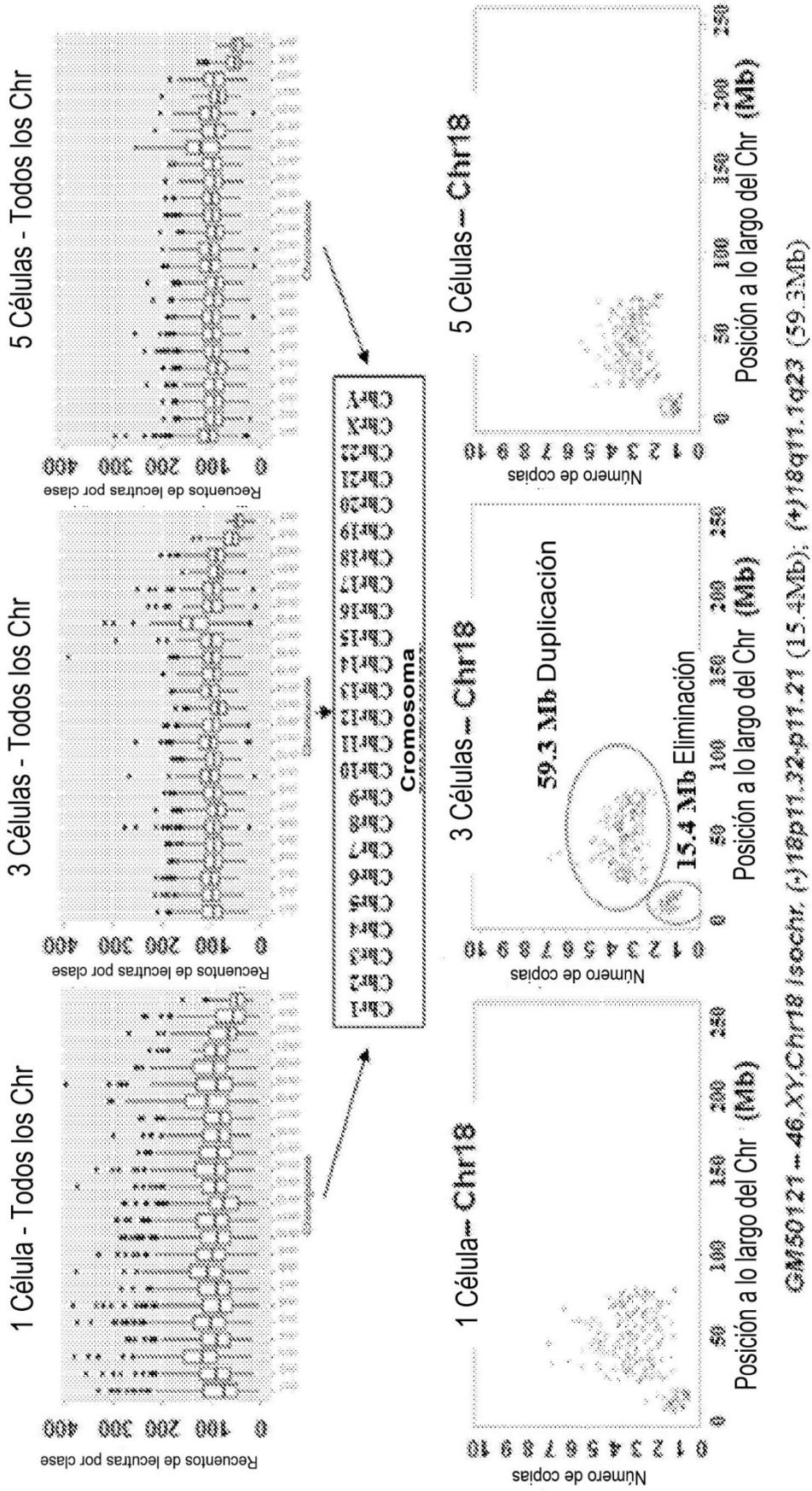
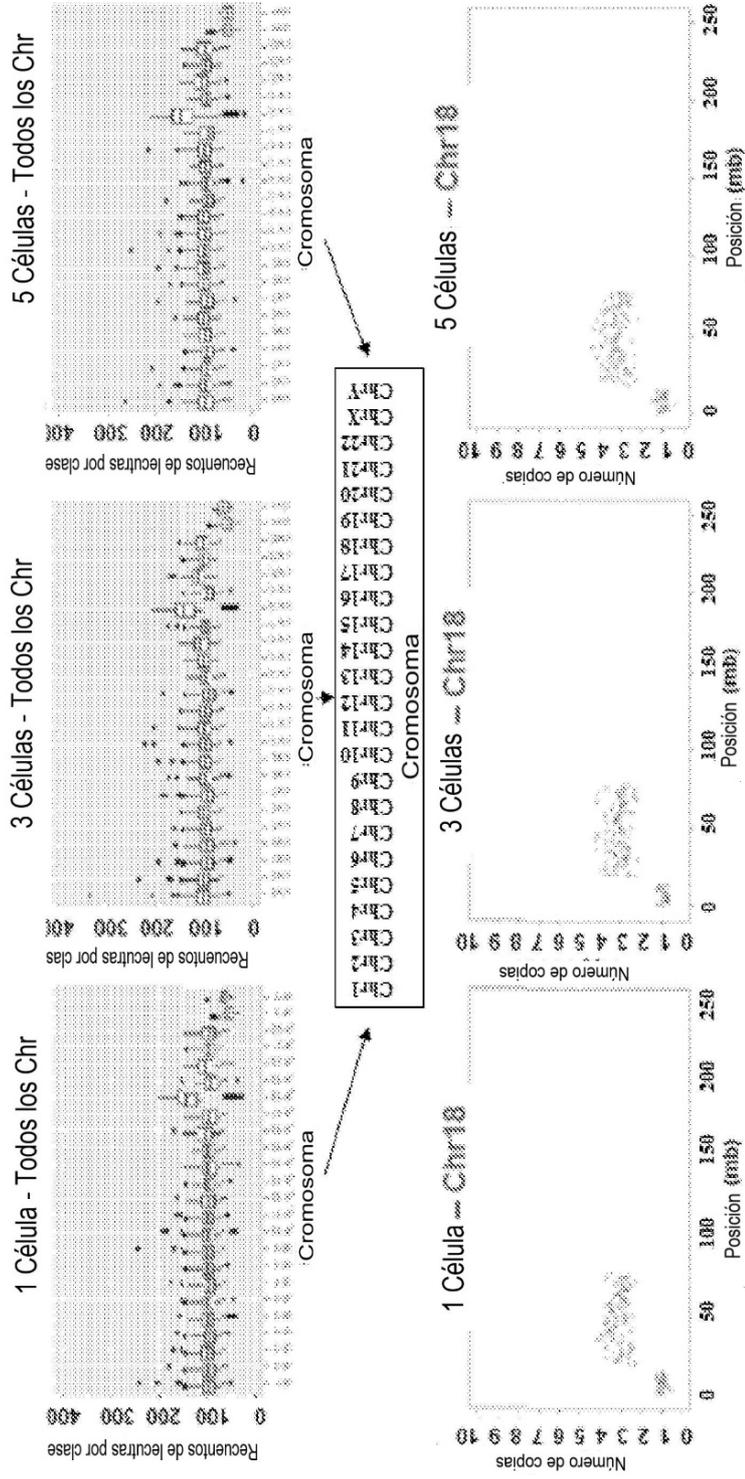


Figura 9A

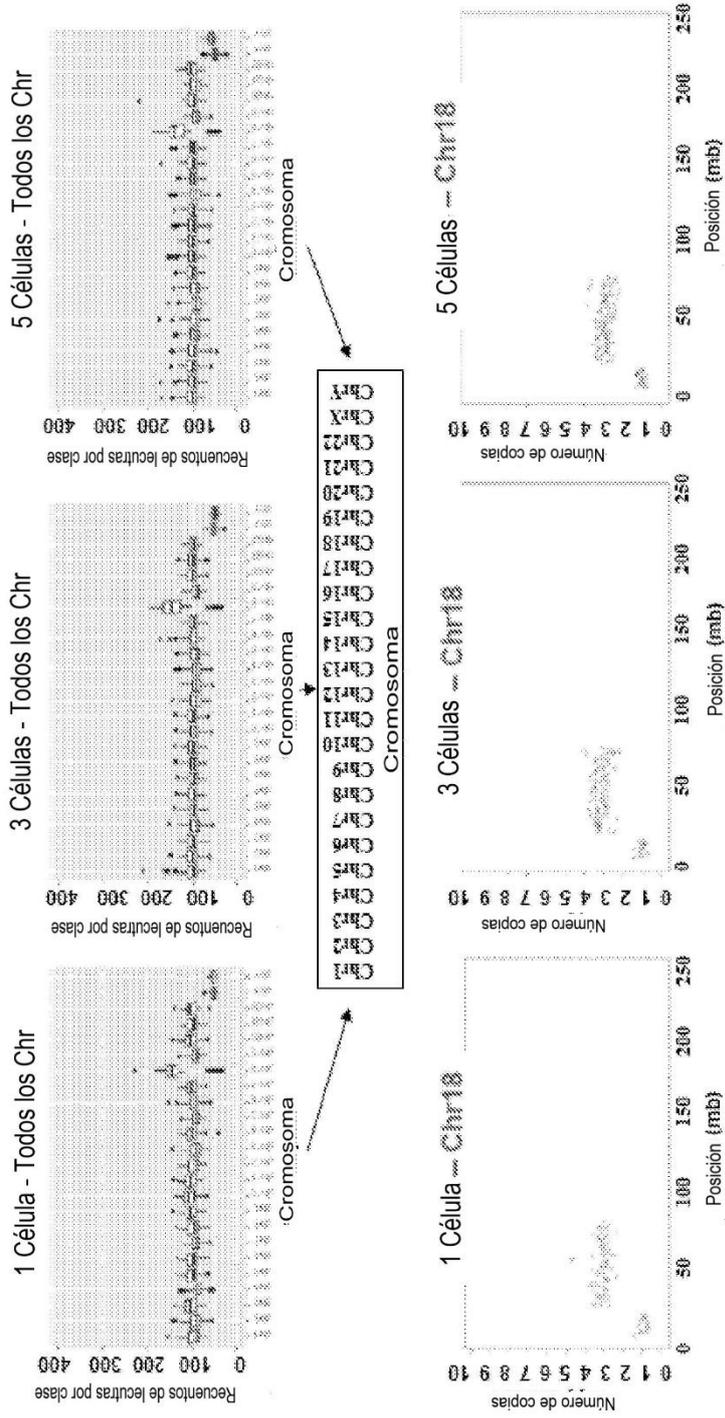
SurePlex (WGA basado en PEP/DOP-PCR) + Preparación de bibliotecas Nextera XT



GM50121 ~ 46,XY,Chr18 Isochr, (-)18p11.32-p11.21 (15.4Mb); (+)18q11.1q23 (59.3Mb)

Figura 9B

Nextera SC



GM50121 - 46,XY,Chr18 isochr. (-)18p11.32-p11.21 (15.4Mb); (+)18q11.1q23 (59.3Mb)

Figura 9C

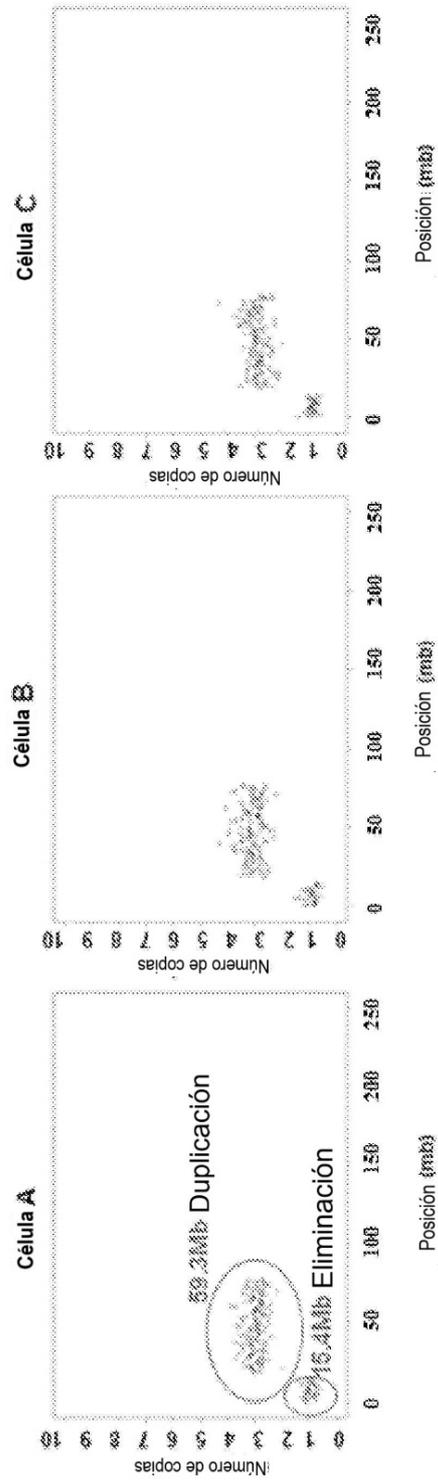


Figura 10A

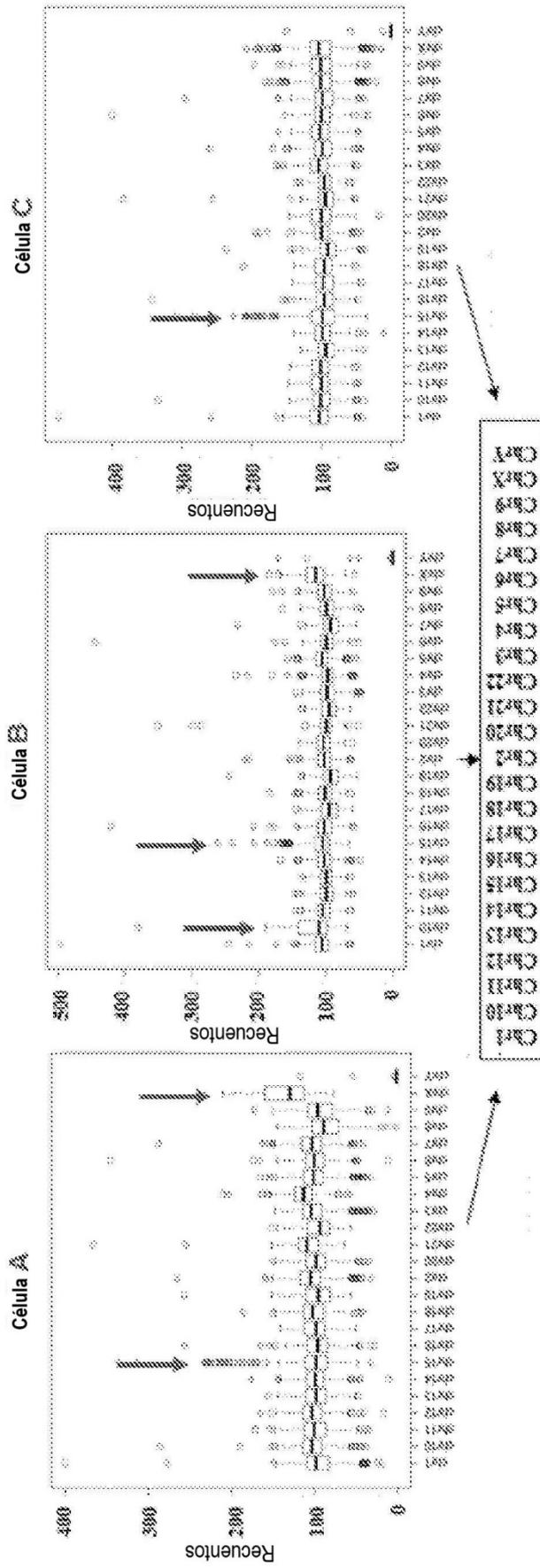


Figura 10B

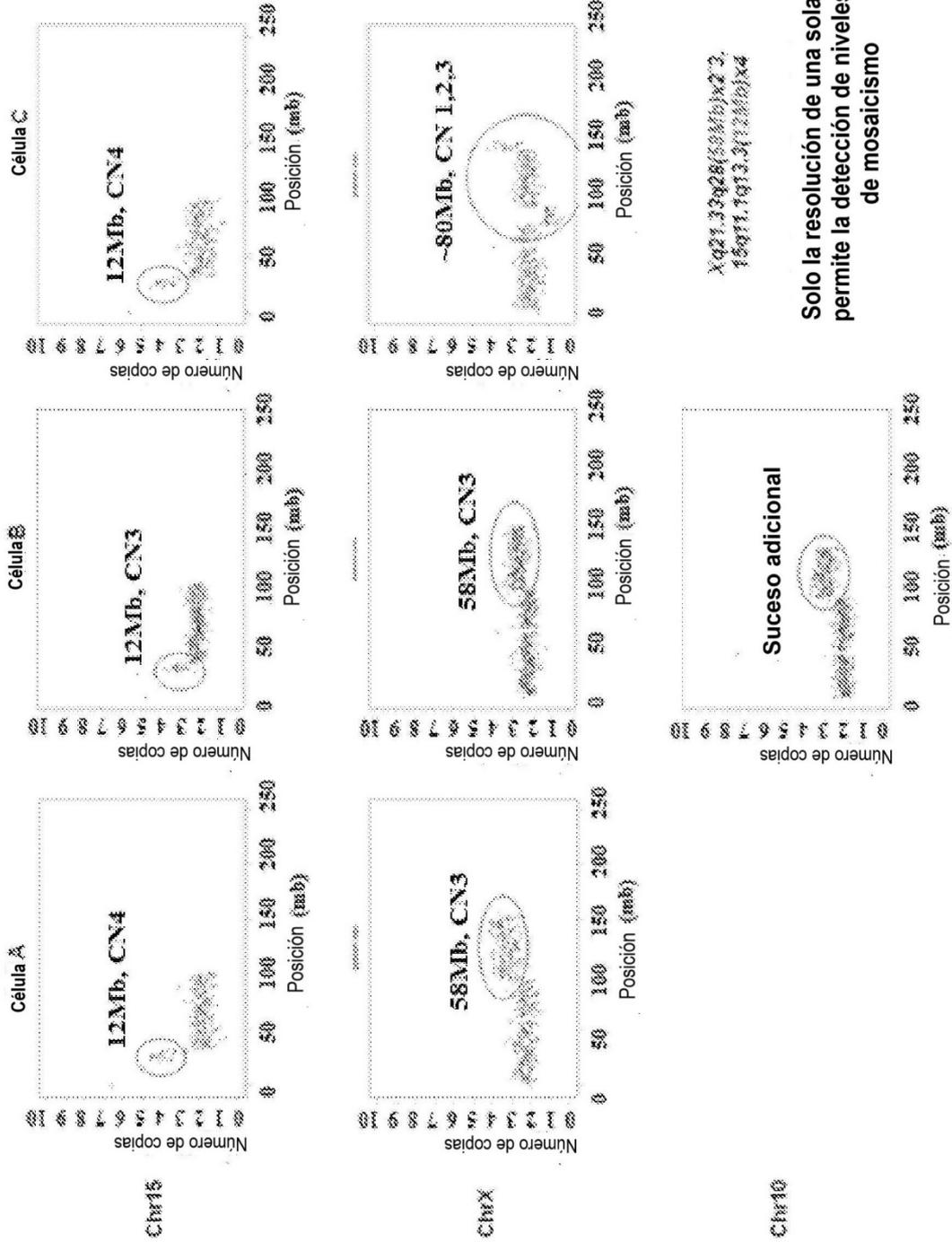


Figura 10C

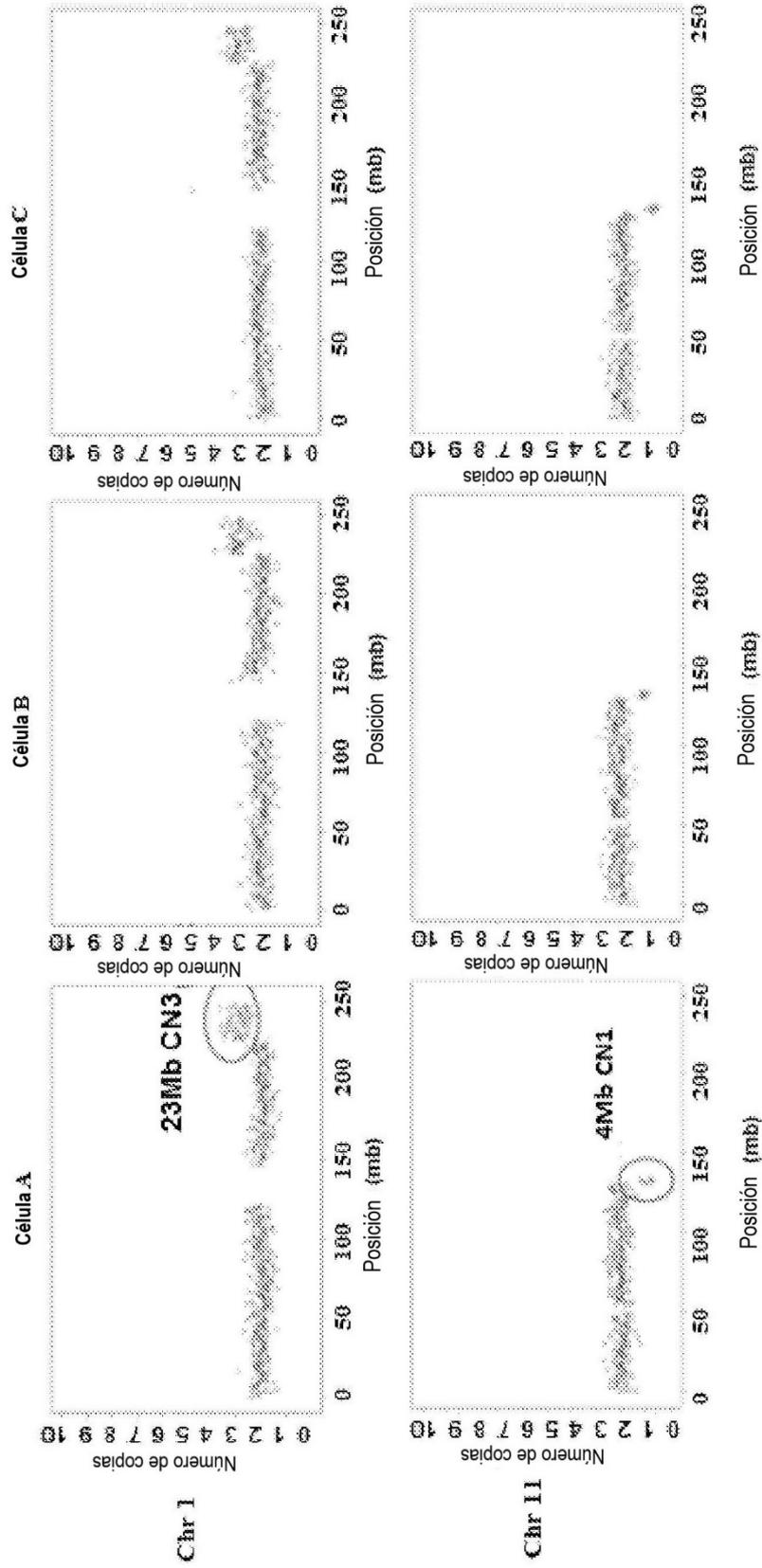


Figura 10D