

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 253**

51 Int. Cl.:

C12Q 1/6869 (2008.01)

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.08.2021 PCT/EP2021/073203**

87 Fecha y número de publicación internacional: **24.02.2022 WO22038291**

96 Fecha de presentación y número de la solicitud europea: **20.08.2021 E 21766455 (6)**

97 Fecha y número de publicación de la concesión europea: **10.04.2024 EP 4200443**

54 Título: **Un método para el aislamiento de roturas de doble hebra**

30 Prioridad:

21.08.2020 GB 202013141

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

29.10.2024

73 Titular/es:

**UNIVERSITY COLLEGE CARDIFF CONSULTANTS LTD (100.0%)
30-36 Newport Road
Cardiff CF24 0DE, GB**

72 Inventor/es:

**REED, SIMON;
DOBBS, FELIX y
VAN EIJK, PATRICK**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

Observaciones:

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

ES 2 984 253 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

Un método para el aislamiento de roturas de doble hebra

5 **Campo de la invención**

10 La invención se refiere a un método para determinar el número y la naturaleza de las roturas de doble hebra (DSB) del ADN en una muestra de ácido nucleico, idealmente ADN genómico (ADNg); a un kit de piezas para realizar el método antes mencionado, que incluye al menos una pluralidad de oligonucleótidos para ligarse a dicha muestra de ácido nucleico; y a oligonucleótidos para uso en dicho kit y dicho método.

Antecedentes de la invención

15 Las roturas en los ácidos nucleicos, en las que se cortan la o las hebras principales del ácido nucleico, son particularmente peligrosas para las células y los organismos debido a que pueden provocar reordenamientos del genoma. Entre tales roturas se incluyen las roturas de doble hebra (DSB), normalmente en el ADN, que son las más peligrosas de todas las lesiones del ADN, ya que comprometen directamente la estabilidad del genoma si no se reparan. Además de provocar la muerte celular y la imposibilidad de repararse fielmente, las DSB pueden provocar carcinogénesis mediante la formación de alteraciones genómicas estructurales, que incluyen deleciones, inserciones, translocaciones de ADN, y eventos de recombinación mitótica en células somáticas. En células sanas, se estima que se forman hasta 50 DSB endógenas por célula durante el ciclo celular. Estas roturas fisiológicas de bajo nivel surgen esporádicamente como resultado de procesos celulares normales, tal como la replicación, transcripción, y formación de bucles de cromatina del ADN, u ocurren en niveles más altos debido a la formación de roturas recurrentes programadas por la célula para facilitar procesos tales como la recombinación V(D)J durante la meiosis. La recombinación V(D)J es una característica definitoria del sistema inmunológico adaptativo. Es un mecanismo único de recombinación genética que ocurre en los linfocitos en desarrollo durante las primeras etapas de maduración de las células B y T. Implica recombinación somática, y da como resultado un repertorio muy diverso de anticuerpos/inmunoglobulinas y receptores de células T (TCR) que se encuentran en las células B y T, respectivamente.

20 Además de las DSB adquiridas a través de procesos celulares normales, una variedad de agentes físicos y químicos exógenos, tales como la radiación ionizante, los fármacos quimioterapéuticos, y más recientemente, las tecnologías de edición del genoma CRISPR, también son potentes inductores de roturas de hebras, particularmente las DSB genómicas.

25 Comprender los procesos que generan roturas en los ácidos nucleicos, particularmente en el genoma, y los mecanismos que los reparan es de importancia central para la medicina genómica. Desde la determinación de las causas y curas del cáncer hasta el desarrollo seguro de técnicas de edición del genoma, la medida precisa y exacta de la frecuencia, posición y causa de las roturas de la doble hebra del ADN en el genoma es primordial. Sin embargo, medir simultáneamente en las células el panorama de todo el genoma de las DSB endógenas y exógenas/inducidas es un desafío, principalmente debido a la gama extremadamente amplia de eventos de rotura esporádicos raros, frente a los recurrentes inducidos con frecuencia.

30 La aparición de la secuenciación de próxima generación (NGS) ha impulsado el desarrollo de numerosos métodos para detectar y medir las secuencias de ADN asociadas con roturas de hebra de ácidos nucleicos, específicamente DSB, a escala genómica. Estos métodos se dividen en términos generales en tres categorías:

- 35 i) marcaje indirecto de roturas usando proteínas como sustitutos para las roturas (por ejemplo, VH2AX ChIP-seq, DISCOVER-seq);
- 40 ii) marcaje indirecto de roturas reparadas (por ejemplo, GUIDE-seq, HTGTS), y
- 45 iii) marcaje directo de extremos rotos no reparados en las células (por ejemplo, BLESS, DSBCapture, END-seq, BLISS).

50 A pesar de las muchas mejoras técnicas incrementales realizadas en cada iteración de estos métodos, todos adolecen de un inconveniente fundamental común; una dependencia de la preparación de la biblioteca de ADN estándar requerida para NGS, después del marcaje y enriquecimiento de las DSB en una muestra mediante PCR antes de la secuenciación. La etapa de amplificación por PCR del proceso introduce un sesgo significativo en las bibliotecas de secuenciación de ADN, lo que da como resultado una representación indirecta y distorsionada del patrón original de las DSB presentes en la muestra. Esto es un fenómeno reconocido que hace imposible determinar la cuantificación de la composición original de las DSB en la muestra. Para muchas aplicaciones de NGS, tal como la secuenciación completa del genoma/exoma, esto puede no representar un problema importante. Pero para la medida cuantitativa de características específicas en todo el genoma, tales como las DSB, el sesgo de la amplificación por PCR introduce altos niveles de ruido en un sistema en el que las señales (DSB) ya son muy bajas.

Para superar esta desventaja, hemos diseñado un nuevo protocolo de preparación de bibliotecas de ADN que evita la necesidad de amplificar la secuencia de rotura mediante PCR antes de que se detecten las DSB, y, además, mejora la señal de las DSB. Al mejorar la relación señal-ruido para la detección de DSB, obtenemos una medida directa de las roturas genómicas en la muestra, en la que una secuencia leída equivale a una rotura, idealmente una DSB, presente en la muestra.

El documento WO2015/134552 A1 describe un método de ligación de adaptadores a los extremos de moléculas de ADN de doble hebra fragmentadas.

Sumario de la invención

Aquí, describimos nuestro nuevo método, denominado INDUCE-seq, para la medida directa de roturas de ácidos nucleicos (típicamente las DSB). Además, demostramos su capacidad para identificar patrones de roturas que pueden caracterizar roturas formadas por una variedad de diferentes causas fisiológicas y también inducidas.

INDUCE-seq puede así detectar, simultáneamente, la presencia de roturas esporádicas de doble hebra de bajo nivel causadas por la transcripción fisiológica y la replicación del ADN, y roturas recurrentes de mayor nivel inducidas por enzimas de restricción o nucleasas de edición del genoma tal como CRISPR-Cas9. Por tanto, INDUCE-seq se puede usar para determinar los orígenes de la formación de roturas del ADN y su mecanismo de reparación, así como para el desarrollo seguro de la edición del genoma.

En un aspecto de la invención, se proporciona un método de preparación de muestras para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean:

- (i) adecuados para unirse a un sustrato que comprende un primer cebador inmovilizado, y
- (ii) adecuados para amplificación, en el que la amplificación comprende el uso de dicho primer cebador inmovilizado y un segundo cebador; comprendiendo el método:
 - a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
 - b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de unirse al primer cebador inmovilizado mediante hibridación;
 - c) fragmentar la pluralidad de ácidos nucleicos; y
 - d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación, y que comprende una secuencia idéntica a una región del segundo cebador.

El segundo cebador puede inmovilizarse al sustrato.

En una realización, la etapa b) es: exponer la pluralidad de ácidos nucleicos a un primer par de adaptadores en condiciones propicias para la ligación, en el que el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y en el que el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación; y la etapa d) es:

exponer la pluralidad de ácidos nucleicos a un segundo par de adaptadores en condiciones propicias para la ligación, en el que el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 5' y comprende una secuencia idéntica a una región de un segundo cebador, y el segundo oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del segundo cebador.

En una realización del método, el segundo par de adaptadores comprende:

un primer oligonucleótido que comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32), y un segundo oligonucleótido que comprende una secuencia que no comprende una secuencia de más de 5, 10, 15, o 20 bases, o que no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); o

5 un primer oligonucleótido que comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34), y un segundo oligonucleótido que comprende una secuencia que no comprende una secuencia de más de 5, 10, o 15 bases, o que no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31).

10 El primer y/o el segundo oligonucleótido del primer par de adaptadores puede comprender una característica de protección de 3' y/o 5'; y/o en el que el primer y/o el segundo oligonucleótido del segundo par de adaptadores comprende una característica de protección de 3' y/o 5'. En una realización, el segundo adaptador no puede ligarse al primer adaptador debido a la presencia de una modificación 3' del primer adaptador.

15 En una realización, el oligonucleótido del segundo adaptador que se puede ligar a un extremo 5' comprende una secuencia idéntica a 5, 10, 15, 20, 21, 24, o más bases de un cebador inmovilizado.

20 El método puede comprender además desnaturalizar la pluralidad de ácidos nucleicos para formar una pluralidad de ácidos nucleicos monocatenarios. El método puede comprender además poner en contacto la pluralidad de ácidos nucleicos con el sustrato que comprende cebadores inmovilizados, en condiciones adecuadas para la hibridación de los cebadores inmovilizados con ácidos nucleicos complementarios. El método puede comprender además obtener información de secuencia para cualesquiera ácidos nucleicos hibridados con el sustrato.

25 En una realización del método, la muestra que comprende una pluralidad de ácidos nucleicos es ADNg.

En algunas realizaciones, las etapas se llevan a cabo en el orden a), b), c), y después d). En otras realizaciones, las etapas se llevan a cabo en el orden a), c), d), y después b); en las que la muestra está expuesta a condiciones capaces de causar o que se sospecha que pueden causar una DSB entre las etapas d) y b).

30 En algunas realizaciones, la muestra se expone a condiciones capaces de provocar una DSB en una característica de interés en la muestra de ácido nucleico.

35 En otro aspecto de la invención se proporciona un kit para la preparación de muestras para identificar las DSB en una muestra de ADNg, en el que el kit comprende pares de oligonucleótidos que están configurados para ser adecuados para actuar como el primer y segundo adaptador cuando se llevan a cabo los métodos de la invención, comprendiendo el kit

40 i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

45 ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

50 En otro aspecto de la invención se proporciona un kit para la preparación de muestras para identificar las DSB en una muestra de ADNg, en el que el kit comprende pares de oligonucleótidos que están configurados para ser adecuados para actuar como el primer y segundo adaptador cuando se llevan a cabo los métodos de la invención, comprendiendo el kit

55 i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

60 ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho

65

oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34); y

en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

El primer oligonucleótido del primer par y/o del segundo par de oligonucleótidos puede comprender un sitio de hibridación al que puede unirse un primer cebador de secuenciación. El primer y segundo oligonucleótidos de la parte i) y/o la parte ii) pueden comprender una característica de índice que es una secuencia particular de nucleótidos que permite determinar el origen de las muestras agrupadas.

El kit puede comprender además al menos un cebador que se une al primer y/o segundo sitio de hibridación con fines de secuenciación. El kit puede comprender además agentes de fragmentación y/o agentes desnaturizantes para fragmentar y/o desnaturizar el ácido nucleico en fragmentos y/o hebras individuales, respectivamente.

En otro aspecto de la invención, se proporciona un adaptador de doble hebra para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, en el que el adaptador de doble hebra está configurado para ser adecuado para actuar como segundo adaptador cuando se llevan a cabo los métodos de la invención, comprendiendo el adaptador de doble hebra: una primera hebra de oligonucleótido que no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

En otro aspecto de la invención, se proporciona un adaptador de doble hebra para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, en el que el adaptador de doble hebra está configurado para ser adecuado para actuar como segundo adaptador cuando se llevan a cabo los métodos de la invención, comprendiendo el adaptador de doble hebra: una primera hebra de oligonucleótido que no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCCGATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

Descripción de los dibujos

Figura 1. Descripción general de INDUCE-seq. El marcaje de rotura *in situ* en células fijadas y permeabilizadas se lleva a cabo ligando un adaptador de secuenciación P5 modificado químicamente de longitud completa a las DSB preparadas en los extremos. Después, el ADN genómico se extrae, se fragmenta, se preparan los extremos y se liga usando un adaptador P7 semifuncional modificado químicamente. Las bibliotecas de ADN resultantes contienen una mezcla de fragmentos funcionales marcados con DSB (P5:P7) y fragmentos de ADN genómico no funcionales (P7:P7). La secuenciación posterior de las bibliotecas INDUCE-seq enriquece los fragmentos marcados con ADN y elimina todo el resto del ADN no funcional. Como la preparación de la biblioteca INDUCE-seq está libre de PCR, cada lectura de secuenciación obtenida equivale a un único extremo de DSB marcado;

Figura 2. Esquema detallado del diseño del adaptador de enriquecimiento de celda de flujo. (a) Estructura de un fragmento de ADNbc ligado a un adaptador completo para secuenciación. Los adaptadores P5 y P7 de 3' se hibridan con la celda de flujo. Los cebadores de secuenciación se unen a las secuencias del cebador de secuenciación de lectura 1 (RD1 SP) y del cebador de secuenciación de lectura 2 (RD2 SP) durante la primera y segunda lectura de secuenciación. Los índices permiten diferenciar diferentes muestras en un conjunto de bibliotecas. (b) Estructura de fragmentos de ADN presentes en una biblioteca INDUCE-seq. Sólo el fragmento ligado a DSB está compuesto por todos los componentes del adaptador necesarios para la secuenciación. (c) La carga de la biblioteca INDUCE-seq en la celda de flujo de secuenciación se enriquecerá con fragmentos ligados a DSB mediante la hibridación de los extremos 3' de las secuencias del adaptador P5. Ningún otro fragmento puede interactuar con la celda de flujo, y será eliminado;

Figura 3. INDUCE-seq demuestra una sensibilidad y un rango dinámico incomparables en comparación con tecnologías de secuenciación de DSB alternativas. (a) INDUCE-seq detecta las DSB inducidas altamente recurrentes y las DSB endógenas únicas simultáneamente con alta resolución. Vista del navegador del genoma de las lecturas de INDUCE-seq cartografiadas a una sección de 10 mb del genoma de las células HEK293T tras la escisión *in situ* con la endonucleasa de restricción HindIII. (Panel superior) Las roturas inducidas por enzimas altamente recurrentes representan la gran mayoría de las lecturas cuando se visualizan en el nivel alto (10 mb, 0-1000 lecturas). (Panel inferior) Un zoom más cercano (resaltado en rosa, 500 kb, 0-20 lecturas) revela roturas endógenas únicas de bajo nivel presentes tanto en la muestra no tratada

como entre las roturas recurrentes inducidas por HindII (resaltado en verde). (b) El cartografiado de lecturas de INDUCE-seq en un sitio diana de HindII demuestra la precisión del cartografiado de la rotura de un solo nucleótido. Esta figura incluye TACTCAAGCTTACCCCTA (SEQ ID NO: 35) y GGGGGGTAAGCTTGAGTA (SEQ ID NO: 43). (c) Cuantificación de las roturas medidas por celda para las muestras tratadas con HindII y de control. INDUCE-seq detecta cuantitativamente roturas por celda en 3 órdenes de magnitud entre muestras. (d y e) Comparación entre INDUCE-seq y DSBCapture en la detección de sitios de restricción escindidos *in vitro* por las enzimas HindII y EcoRV. (d) Una mayor proporción de lecturas secuenciadas y alineadas con el genoma se cartografiaron en sitios de restricción usando INDUCE-seq. (e) Usando 800 veces menos células, INDUCE-seq identifica una proporción similar de sitios de restricción HindII (92,7 %) a la identificada por DSBCapture para EcoRV (93,7 %). (f) El rango dinámico de detección de DSB inducida usando INDUCE-seq. Además de las roturas identificadas en las secuencias en la diana de HindII (AAGCTT), también se identificaron múltiples sitios fuera de la diana no coincidentes de 1 y 2 pb. INDUCE-seq midió eventos de roturas inducidas que abarcan 8 órdenes de magnitud, desde ~150 millones de roturas identificadas en los sitios en la diana de HindII hasta 5 roturas identificadas en los sitios más raros fuera de la diana. (g) Comparación entre INDUCE-seq, DSBCapture y BLISS en la detección de roturas inducidas por AsiSI en células DiVA vivas. El número de lecturas secuenciadas (panel superior) se compara con el número de sitios AsiSI identificados para cada experimento (panel inferior). INDUCE-seq detecta con sensibilidad la mayor cantidad de sitios AsiSI usando 40 veces menos lecturas que DSBCapture y 23 veces menos lecturas que BLISS;

Figura 4. INDUCE-seq descubre y cuantifica con sensibilidad las DSB dentro y fuera de la diana inducidas por CRISPR/Cas9. (a) Secuencias fuera de la diana y número de roturas identificadas mediante INDUCE-seq para el sgRNA de EMX1. Esta figura incluye GAGTCCGAGCAGAAGAAGAANGG (SEQ ID NO: 44). (b) INDUCE-seq revela la cinética de la formación de las DSB inducidas por EMX1 en una población celular. La cuantificación del número de roturas detectadas por millón de lecturas para cada muestra reveló una alta actividad de Cas9 tanto dentro como fuera de la diana inmediatamente después de la nucleofección celular. (c) La comparación entre sitios fuera de la diana identificados por INDUCE-seq con los métodos *in vitro* establecidos CIRCLE-seq y Digenome-seq, además de los métodos basados en células GUIDE-seq, BLISS y HTGTS. INDUCE-seq detecta muchos sitios fuera de la diana que anteriormente sólo eran identificables mediante enfoques *in vitro*. Se identificaron sustancialmente más sitios fuera de la diana que con cualquiera de los métodos celulares actuales. INDUCE-seq también identifica múltiples sitios fuera de la diana no detectados por ningún otro método. (d) Secuenciación de amplicones para medir la frecuencia de indel en los sitios fuera de la diana identificados por INDUCE-seq. La secuenciación de amplicones sólo pudo identificar 4 de los 60 sitios fuera de la diana identificados por INDUCE-seq, y está limitada por la tasa de descubrimiento de falsos indel de fondo del 0,1 %. Estos hallazgos son consistentes con estudios previos que midieron las frecuencias de indel fuera de la diana de EMX1 48 horas después de la nucleofección con RNP de EMX1;

Figura 5. Comparación de INDUCE-seq y los flujos de trabajo de cartografiado de las DSB actuales. (a) Descripción general del flujo de trabajo de INDUCE-seq. La secuenciación de las bibliotecas de INDUCE-seq genera un resultado cuantitativo en el que una lectura equivale a una rotura. (b) Descripción general de los flujos de trabajo de DSBCapture, BLISS y END-seq. La secuenciación siguiendo la construcción de biblioteca estándar genera una lectura de salida no equivalente a una sola DSB;

Figura 6. Comparación entre el número de lecturas secuenciadas y el número de DSB definidas para las bibliotecas de INDUCE-seq y BLISS NGS. (a) Gráfico de dispersión del número de lecturas de INDUCE-seq secuenciadas y el número de roturas definidas a partir de experimentos individuales de INDUCE-seq. (b) Diagrama de dispersión que muestra el número de lecturas de BLISS secuenciadas con el prefijo de código de barras de lectura correcta 1 (R1) y el número de roturas definidas después de la eliminación de duplicados mediante la corrección UMI, de experimentos BLISS individuales;

Figura 7. Vista del navegador del genoma de los puntos de acceso de DSB en células HEK293, (a) vista de 11 kb de un punto de acceso de DSB ch17. Las flechas moradas representan los extremos de DSB marcados en el lado derecho (+ hebra), y las flechas azules representan los extremos de DSB marcados en el lado izquierdo (- hebra). Las DSB recurrentes están distribuidas uniformemente en toda la región del punto de acceso. (b) Vista de 5 kb de un punto de acceso de DSB chr11. Las DSB recurrentes se pueden detectar en diferentes posiciones de las hebras más y menos;

Figura 8. Esquema del proceso de descubrimiento fuera de la diana usado para INDUCE-seq. La secuencia en la Fig. 8 es SEQ ID NO: 44;

Figura 9. El descubrimiento de CRISPR fuera de la diana usando INDUCE-seq es altamente reproducible. (a) Comparación entre el número de sitios fuera de la diana de EMX1 detectados en los experimentos de curso temporal r1 y r2. (b) Diagrama de dispersión que muestra el número de roturas encontrado en los sitios de CRISPR fuera de la diana identificados en ambos experimentos independientes;

Figura 10. Diagramas de Venn que muestran la intersección de los sitios fuera de la diana identificados por INDUCE-seq, CIRCLE-seq, GUIDE-seq y BLISS, (a y b) Superposiciones calculadas para muestras de 0 h a 30 h de forma aislada de los experimentos independientes r1 (a) y r2 (b). (c y d) Las superposiciones combinadas de todos los puntos de tiempo para los conjuntos r1 (c) y r2 (d). (e) Superposiciones calculadas entre métodos cuando se combinan todas las muestras de INDUCE-seq;

Figura 11. El patrón de DSB en sitios dentro y fuera de la diana inducidos por CRISPR se correlaciona con el resultado de la edición. Pistas de cobertura del EMX1 en la diana (a) y los fuera de la diana de clasificación superior (las secuencias mostradas en orden de arriba a abajo son SEQ ID NO: 45-65), OT-1 (las secuencias mostradas en orden de arriba a abajo son SEQ ID NO: 66-87) (b) y OT-2 (las secuencias mostradas en orden de arriba a abajo son SEQ ID NO: 88-105) (c), que abarca 180 pb. Una vista de cerca de la región de 40 pb que rodea cada sitio diana muestra un patrón de escisión distintivo de 1 pb que sobresale en lugar de la DSB roma inducida por Cas9 habitual. Los espectros de indel correspondientes en cada sitio muestran la posición de las mutaciones en relación con los sitios de rotura observados;

Figura 12. Realización ejemplar que indica modificaciones químicas ejemplares.

Figura 13. Realización ejemplar en la que el adaptador semifuncional se liga antes que el adaptador completamente funcional. Esta figura ilustra una realización para la detección de DSB que se inducen artificialmente. Por ejemplo, para la detección de DSB fuera de la diana causadas por nucleasas.

Figura 14. Flujo de trabajo ejemplar para aplicar los conceptos descritos aquí a la tecnología de secuenciación de perlas. Descrito como referencia.

Figura 15. Adaptadores ejemplares para uso con tecnología de secuenciación de perlas. Descrito como referencia.

Figura 16. Método ejemplar para detectar una lesión de ADN que afecta a una hebra. Este ejemplo implica la ligación del adaptador semifuncional antes del adaptador completamente funcional. Descrito como referencia.

Figura 17. Realización ejemplar que muestra la detección *in vitro* de edición de bases de CRISPR - editor de bases de citosina (CBE).

Figura 18. Realización ejemplar que muestra la detección *in vitro* de edición de bases CRISPR - editor de bases de adenina (ABE).

Tabla 1. Secuencias adaptadoras de ejemplo según una realización preferida de la invención.

Descripción detallada

En una realización de la invención, se proporciona un método para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico que comprende

- i) exponer una muestra de ácido nucleico que se sospecha que contiene las DSB, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par y comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una segunda hebra de dichas DSB; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';
- ii) fragmentar la muestra de ácido nucleico, idealmente ADN, en fragmentos;
- iii) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicho fragmento, idealmente en un sitio remoto de la unión de dicho oligonucleótido de dicho primer par, y un sitio de hibridación (RD2 SP) al que puede unirse un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para la unión a una segunda hebra de dichos fragmentos o dichas DSB, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';
- iv) desnaturalizar los fragmentos para proporcionar ácidos nucleicos individuales;

5 v) separar las hebras de la parte iv) en dos grupos: grupo A: aquellos fragmentos que tienen ligado, en un primer extremo, el primer sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte i), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte iii); y grupo B: aquellos fragmentos que no tienen ligados en un primer extremo el sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte i), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte iii); y

10 vi) secuenciar las hebras del grupo A usando cebadores que se unen al primer y/o segundo sitio de hibridación, en el que cada secuencia es equivalente a una rotura, típicamente una DSB. Opcionalmente, además, el número y la naturaleza de las eliminaciones de pares de bases se pueden determinar comparando cada secuencia con un genoma representativo de dicha especie de la que se tomó la muestra.

15 No es necesario realizar las etapas del método de la invención al mismo tiempo o en un solo lugar. Por ejemplo, en un aspecto, el método de la invención permite la preparación de una muestra de ácido nucleico que se ha marcado de una manera que permitiría la selección y el aislamiento de las DSB. Como tal, en una realización de la invención, se proporciona un método de preparación de muestras para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, que comprende

20 i) exponer una muestra de ácido nucleico que se sospecha que contiene las DSB, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par y comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una segunda hebra de dichas DSB; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

30 ii) fragmentar la muestra de ácido nucleico, idealmente ADNg, en fragmentos; y

35 iii) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicho fragmento, opcionalmente en un sitio remoto de la unión de dicho oligonucleótido de dicho primer par, y un sitio de hibridación (RD2 SP) al que puede unirse un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para la unión a una segunda hebra de dicho fragmento, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'.

45 Alternativamente, los métodos de la invención se pueden llevar a cabo hasta que las hebras se separen en el grupo A y el grupo B, y la secuenciación se puede llevar a cabo por separado. Por tanto, el método permite la preparación de una muestra de ácido nucleico en la que se han seleccionado y aislado cualesquiera DSB. Como tal, en una realización, se proporciona un método de preparación de muestras para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, que comprende

50 i) exponer una muestra de ácido nucleico que se sospecha que contiene las DSB, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par y comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una segunda hebra de dichas DSB; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

55 ii) fragmentar la muestra de ácido nucleico, idealmente ADNg, en fragmentos;

60 iii) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicho fragmento, opcionalmente en un sitio remoto de la unión de dicho oligonucleótido de dicho primer par, y un sitio de hibridación (RD2 SP) al que puede unirse un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para la unión a una segunda hebra de dicho fragmento, una secuencia complementaria a dicho sitio de hibridación, y una secuencia

65

adicional, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

5 iv) desnaturalizar los fragmentos para proporcionar ácidos nucleicos individuales; y

5 v) separar las hebras de la parte iv) en dos grupos: grupo A: aquellos fragmentos que tienen ligado en un primer extremo el primer sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte i), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte iii); y grupo B: aquellos fragmentos que no tienen ligados en un primer extremo el sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte i), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte iii).

15 En una realización de la invención, se proporciona un método para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico que comprende

20 i) exponer una muestra de ácido nucleico que se sospecha que contiene las DSB, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

25 ii) fragmentar la muestra de ácido nucleico, por ejemplo ADN_g, en fragmentos;

30 iii) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, uno de los cuales es en parte complementario a un segundo oligonucleótido del segundo par; y un segundo oligonucleótido que comprende una característica de unión 3' para la unión a una segunda hebra de dichos fragmentos, y opcionalmente una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

35 iv) desnaturalizar los fragmentos para proporcionar ácidos nucleicos individuales;

40 v) separar las hebras de la parte iv) en dos grupos: grupo A: aquellos fragmentos que tienen ligada en un primer extremo la secuencia de unión proporcionada por el oligonucleótido de la parte i), y opcionalmente en otro extremo la secuencia de unión para permitir la amplificación del puente proporcionada por el oligonucleótido de la parte iii), y grupo B: aquellos fragmentos que no tienen ligada en un primer extremo la secuencia de unión proporcionada por el oligonucleótido de la parte i), y opcionalmente en otro extremo la secuencia de unión para permitir la amplificación del puente proporcionada por el oligonucleótido de la parte iii); y

45 vi) secuenciar las hebras del grupo A, en el que cada secuencia es equivalente a una rotura, típicamente una DSB, y opcionalmente además en el que el número y la naturaleza de las deleciones de pares de bases se pueden determinar comparando cada secuencia con un genoma representativo de dicha especie de la cual se tomó la muestra. Como se analizó anteriormente, el método puede comprender las etapas i), ii) y iii); o i), ii), iii), iv), y v); o todos las etapas.

En realizaciones particulares, el oligonucleótido del segundo par que comprende una característica de unión 5' no comprende una secuencia de unión para separar dicho ácido nucleico fragmentado.

50 En realizaciones de la invención, los oligonucleótidos de la parte i) y la parte iii) pueden intercambiarse, por lo cual el ácido nucleico se expone primero a los oligonucleótidos de la parte iii) y después se expone a los oligonucleótidos de la parte i).

55 En una realización de la invención, los oligonucleótidos de la parte i) y la parte iii) pueden intercambiarse, por lo cual el ácido nucleico se expone primero a los oligonucleótidos de la parte iii), y tras la fragmentación, los fragmentos de ácido nucleico se exponen después a los oligonucleótidos de la parte i).

60 Las realizaciones que implican el intercambio de la parte i) y la parte iii) son particularmente relevantes para los métodos para la detección de las DSB que se han inducido. En tales realizaciones, la muestra se puede fragmentar y después se puede realizar la etapa iii). Posteriormente, la muestra puede tratarse para inducir potencialmente una DSB y, tras la inducción, puede realizarse la etapa i). Las realizaciones en las que se introduce una DSB se analizan más detalladamente aquí, e incluyen realizaciones para la detección de efectos fuera de la diana de nucleasas, y similares. Las figuras 13, 17 y 18 ilustran realizaciones particulares.

65 En el método de la invención, cualquiera de dicho par de oligonucleótidos puede ser, o puede conocerse como, adaptadores.

La referencia aquí a adaptadores (o adaptadores) es una referencia a un conector en ingeniería genética, y es un oligonucleótido que puede ligarse a los extremos de otras moléculas de ADN. Los adaptadores pueden ser de doble hebra. Los adaptadores de doble hebra se pueden sintetizar para que tengan extremos romos en ambos terminales, o para que tengan un extremo adhesivo en un extremo y un extremo romo en el otro. El adaptador puede ser corto, y se puede sintetizar químicamente.

El primer par de oligonucleótidos puede configurarse para permitir la unión a un sustrato mediante hibridación con un oligonucleótido inmovilizado a dicho sustrato, en el que el oligonucleótido inmovilizado está orientado de manera que el extremo 5' es proximal y el extremo 3' es distal al punto de inmovilización. Por ejemplo, el oligonucleótido del primer par que está ligado al extremo 3' de una hebra de una DSB (es decir, el oligonucleótido que comprende una característica de unión 5') puede ser al menos parcialmente complementario a un oligonucleótido inmovilizado. El alcance de la complementariedad puede permitir la unión al oligonucleótido inmovilizado mediante hibridación. El segundo par de oligonucleótidos puede configurarse para no permitir la unión a un sustrato mediante hibridación con un oligonucleótido inmovilizado a dicho sustrato, en el que el oligonucleótido inmovilizado está orientado de manera que el extremo 5' es proximal y el extremo 3' es distal al punto de inmovilización. Por ejemplo, el oligonucleótido del segundo par que está ligado al extremo 3' de una hebra de un sitio fragmentado (es decir, el oligonucleótido que comprendía una característica de unión 5') puede ser insuficientemente complementario a cualquier oligonucleótido inmovilizado para poder unirse mediante hibridación. En algunas realizaciones, por ejemplo para permitir la amplificación del puente posterior, el oligonucleótido del segundo par que está unido al extremo 5' de una hebra de un sitio fragmentado (es decir, el oligonucleótido que comprende una característica de unión 3') puede ser una secuencia al menos parcialmente idéntica a un oligonucleótido inmovilizado. En otras realizaciones, por ejemplo para permitir la amplificación en emulsión de perlas, se pueden usar otras disposiciones (por ejemplo, véase la Figura 14).

En una realización, se proporciona un método de preparación de muestras para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende oligonucleótidos inmovilizados, comprendiendo el método

i) exponer una muestra de ácido nucleico que se sospecha que contiene las DSB, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB, en el que la secuencia de unión es al menos parcialmente complementaria a un oligonucleótido inmovilizado; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par y comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una segunda hebra de dichas DSB; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

ii) fragmentar la muestra de ácido nucleico, idealmente ADN, en fragmentos; y

iii) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicho fragmento, idealmente en un sitio remoto de la unión de dicho oligonucleótido de dicho primer par, un sitio de hibridación (RD2 SP) al que puede unirse un segundo cebador de secuenciación, y en el que el primer oligonucleótido no comprende una secuencia capaz de hibridarse con un oligonucleótido inmovilizado; y un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para la unión a una segunda hebra de dicho fragmento, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'.

Un ácido nucleico asociado a DSB es un ácido nucleico ubicado en un lado de una DSB. Por tanto, la secuenciación de un ácido nucleico asociado a DSB permite identificar la ubicación de una DSB, por ejemplo en un genoma.

En una realización, los métodos de la invención están diseñados para uso con los adaptadores Illumina P5 y P7. Así, en una realización particular, el segundo par de oligonucleótidos no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30). En otra realización, el segundo par de oligonucleótidos no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTGCCTGATCATT (SEQ ID NO: 31).

Así, en una realización particular, la etapa iii) es:

exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicho fragmento, idealmente en un sitio remoto de la unión de dicho oligonucleótido de

dicho primer par, y un sitio de hibridación (RD2 SP) al que puede unirse un segundo cebador de secuenciación; y

5 un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para la unión a una segunda hebra de dichas DSB, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'; y en el que el segundo par de oligonucleótidos no comprende una secuencia de más de 5, 10, 15, 20, o 24 bases de SEQ ID NO: 30 y/o no comprende una secuencia de más de 5, 10, 15, o 20 bases de SEQ ID NO: 31.

15 En el método preferido de la invención, dicha característica de unión 5' o 3' de dicho par de oligonucleótidos comprende uno de los siguientes: un grupo fosfato; una "cola T" de trifosfato, preferiblemente una "cola T" de trifosfato de desoxitimidina; una "cola A" de trifosfato, preferiblemente una "cola A" de trifosfato de desoxiadenosina; al menos un N nucleótido aleatorio, preferiblemente una pluralidad de N nucleótidos, o cualquier otro grupo de unión conocido para permitir la unión de dicho adaptador a dicha DSB.

20 La muestra de ácido nucleico puede ser cualquier muestra de ADN capaz de comprender las DSB. En una realización preferida de la invención, dicha muestra de ácido nucleico es ADNg.

Aún más idealmente, dicha característica de unión 5' de dicho primer oligonucleótido de la parte i) es un grupo fosfato, y dicha característica de unión 3' de dicho segundo oligonucleótido de la parte i) es una cola de trifosfato.

25 En una realización preferida de la invención, dicha característica de protección de 5' y/o 3' de dicho primer par de oligonucleótidos comprende una característica que proporciona resistencia a una cualquiera o más de las siguientes: actividad de fosforilación, actividad de fosfatasa, actividad de transferasa terminal, hibridación de ácido nucleico, actividad de endonucleasa, actividad de exonucleasa, actividad de ligasa, actividad de polimerasa, y unión a proteínas. Esto se puede lograr por cualquier medio conocido por los expertos en la técnica, tales como, pero sin limitarse a, enlaces de fosforotioato, espaciadores de fosforoamidita, grupos fosfato, grupos 2'-O-metilo, modificaciones desoxi y didesoxi-T invertidas, bases de ácidos nucleicos bloqueados, didesoxinucleótidos, o similares. En la tabla 2 se muestran ejemplos de la actividad que proporcionan estas características.

35 Preferiblemente, dicho primer oligonucleótido de la parte i) comprende una característica de protección de 3' que proporciona resistencia a la actividad exonucleasa, tal como un enlace de fosforotioato.

40 Adicional o alternativamente, dicha característica de protección 3' también proporciona resistencia a la actividad de ligasa y/o actividad de polimerasa, idealmente, actividad de polimerasa 5'>3', y es, por ejemplo, un didesoxinucleótido o un bloque físico, idealmente en forma de fosforamidita, en particular una fosforamidita espaciadora C3 (3SpC3), o cualquier otra característica de protección conocida por los expertos en la técnica, tal como las de la Tabla 2, que proporciona resistencia a la actividad de exonucleasa y/o actividad de ligasa y/o actividad de polimerasa.

45 Aún más preferiblemente, dichos primer y segundo oligonucleótidos de la parte i) también comprenden una característica de índice que es una secuencia particular de nucleótidos (por ejemplo, GATCT) que permite determinar el origen de las bibliotecas de secuenciación reunidas; en otras palabras, permite la desmultiplexación de las bibliotecas de secuenciación reunidas. Idealmente, la característica de índice está situada entre dicho sitio de hibridación y dicha secuencia de unión.

50 Lo más preferible, dicho primer oligonucleótido de la parte i), que lee de 5' a 3', comprende una característica de unión 5', y después, opcionalmente, una característica de protección, idealmente la característica de unión es un grupo fosfato, un sitio de hibridación (RD1 SP) al cual se puede unir un primer cebador de secuenciación, una secuencia índice, una secuencia de unión para separar dicha DSB de un conjunto de DSB, y una característica de protección y/o de unión 3'. Preferiblemente, dicha característica de protección proporciona resistencia a una cualquiera o más de las siguientes: actividad de exonucleasa, actividad de ligasa y/o actividad de polimerasa.

55 Lo más preferible, dicho segundo oligonucleótido de la parte i), que lee de 3' a 5', comprende una característica de unión 3', y después, opcionalmente, una característica de protección, una secuencia de hibridación (RD2 SP) a la cual se puede unir un primer cebador de secuenciación, una secuencia índice, una secuencia de unión para separar dicha DSB de un conjunto de DSB, y una característica de protección y/o de unión 5'. Preferiblemente, dichas características de protección proporcionan resistencia a una cualquiera o más de las siguientes: actividad de exonucleasa, actividad de ligasa y/o actividad de polimerasa. Idealmente, la característica de unión es 3'.

65 En una realización preferida de la invención, dicho primer oligonucleótido de la parte i) es uno de un primer par de oligonucleótidos, y el segundo oligonucleótido de este primer par de oligonucleótidos es complementario a dicho primer oligonucleótido y comprende una característica de protección de 5' y 3', preferiblemente, que proporciona resistencia a una cualquiera o más de las siguientes: actividad de exonucleasa, actividad de ligasa y/o actividad de polimerasa.

Por tanto, en ciertas realizaciones, falta un grupo fosfato 5' o 3' en el oligonucleótido usado para llevar a cabo la invención.

En una realización preferida de la invención, dicho primer oligonucleótido de la parte i) contiene tanto el sitio/(secuencia) de hibridación que se usa para secuenciar la DSB ligada como la secuencia de unión que se usa para separar dicha DSB de un conjunto de DSB. Sin embargo, los expertos en la técnica apreciarán que es posible que el segundo oligonucleótido de la parte i) contenga tanto el sitio/(secuencia) de hibridación que se usa para secuenciar la DSB ligada como la secuencia de unión que se usa para separar dicha DSB. de un conjunto de DSB. Esto se debe a que tanto la secuenciación como la separación estarán determinadas por la naturaleza del cebador usado para secuenciar y la naturaleza del oligonucleótido usado para separar. Lo más típico, la orientación de todos los oligonucleótidos, es decir, la orientación de los sitios de hibridación (RD1 SP y RD2 SP) a los que se pueden unir al menos un primer y/o segundo cebadores de secuenciación, y la secuencia de unión para separar dicha DSB de un conjunto de DSB, y la secuencia adicional, opcionalmente, para permitir la amplificación del puente, es tal que, cuando se lleva a cabo la separación de la parte v), sólo se pueden extraer hebras del Grupo A usando la secuencia de unión para separar dicha DSB de un conjunto de DSB. Además, esta orientación también es tal que cuando se lleva a cabo la secuenciación de la parte vi) sólo se pueden unir hebras del Grupo A (cuando está presente la secuencia de amplificación del puente).

Más idealmente aún, el primer o el segundo oligonucleótido del primer par de oligonucleótidos de la parte i) comprende dos características de protección terminales diferentes.

Aún más preferentemente, el segundo oligonucleótido de este primer par de oligonucleótidos de la parte i) comprende una 'cola T' de trifosfato de 3' desoxitimidina, para proporcionar un sustrato para la ligación a fragmentos de ADN de 'cola A', e idealmente también un enlace de fosforotioato mediante el cual se confiere resistencia a la actividad de exonucleasa.

En aún otro método preferido de la invención, dicha característica de unión 5' de dicho primer oligonucleótido del segundo par de oligonucleótidos de la parte iii) es un grupo fosfato, y dicha característica de unión 3' de dicho segundo oligonucleótido del segundo par de oligonucleótidos de la parte iii) es una cola de trifosfato.

En una realización preferida de la invención, dicha característica de protección de 5' y/o 3' de dicho segundo par de oligonucleótidos comprende una característica que proporciona resistencia a uno cualquiera o más de los siguientes: actividad de fosforilación, actividad de fosfatasa, actividad de transferasa terminal, hibridación de ácido nucleico, actividad de endonucleasa, actividad de exonucleasa, actividad de ligasa, actividad de polimerasa, y unión a proteínas. Esto se puede lograr por cualquier medio conocido por los expertos en la técnica, tales como, pero sin limitarse a, enlaces de fosforotioato, espaciadores de fosfoamidita, grupos fosfato, grupos 2'-O-metilo, modificaciones desoxi y didesoxi-T invertidas, bases de ácidos nucleicos bloqueados, didesoxinucleótidos, o similares. En la tabla 2 se muestran ejemplos de la actividad que proporcionan estas características.

Preferiblemente, dicho primer oligonucleótido de la parte iii) también comprende una característica de protección de 3' que proporciona resistencia a la actividad exonucleasa, tal como un espaciador de fosforamidita. Adicional o alternativamente, dicha característica de protección 3' también proporciona resistencia a la actividad de ligasa y/o actividad de polimerasa, idealmente, actividad de polimerasa 5'>3', y es, por ejemplo, un didesoxinucleótido o un bloque físico, idealmente en forma de fosforamidita, en particular una fosforamidita espaciadora C3 (3SpC3), o cualquier otra característica de protección conocida por los expertos en la técnica, tal como las de la Tabla 2, que proporciona resistencia a la actividad de exonucleasa y/o actividad de ligasa y/o actividad de polimerasa.

Aún más preferiblemente, dichos primer y segundo oligonucleótidos de la parte iii) también comprenden una característica de índice que es una secuencia particular de nucleótidos (por ejemplo, GATCT) que permite determinar el origen de las bibliotecas de secuenciación reunidas; en otras palabras, permite la desmultiplexación de las bibliotecas de secuenciación reunidas. Idealmente, la característica de índice está situada entre dicho sitio de hibridación y, cuando esté presente, dicha secuencia adicional.

Más idealmente aún, el primer o el segundo oligonucleótido de este segundo par de oligonucleótidos de la parte iii) comprende dos características de protección terminales diferentes.

Más preferiblemente, dicho segundo oligonucleótido de la parte iii), que se lee de 5' a 3', comprende una característica de unión y/o protección de 5', una secuencia adicional para, opcionalmente, permitir la amplificación del puente, una secuencia de índice, un sitio de hibridación (RD2 SP). al que se puede unir un cebador de secuenciación, y una característica de y/o unión protección de 3'. Preferiblemente, una o ambas características de protección proporcionan resistencia a una cualquiera o más de las siguientes: actividad de exonucleasa, actividad de ligasa y/o actividad de polimerasa.

En una realización preferida de la invención, dicho segundo oligonucleótido de la parte iii) es complementario a dicho primer oligonucleótido de este par de oligonucleótidos, y comprende una característica de protección de 5' y 3', preferiblemente, que proporciona resistencia a una cualquiera o más de las siguientes: actividad de exonucleasa,

actividad de ligasa y/o actividad de polimerasa. Por lo tanto, en ciertas realizaciones, el oligonucleótido usado para llevar a cabo la invención carece de un grupo fosfato 5' o 3'.

5 Aún más preferiblemente, este segundo oligonucleótido de la parte iii) comprende una 'cola T' de trifosfato de 3' desoxitimidina, para proporcionar un sustrato para la ligación a fragmentos de ADN con 'cola A', idealmente con un enlace de fosforotioato mediante el cual se confiere resistencia a la actividad de exonucleasa.

10 En una realización preferida de la invención, dicha ligación en la parte i) se produce *in situ* o *in vitro* usando una suspensión celular o tisular, y por tanto, se produce en la célula intacta. Aún más preferiblemente, esta etapa se facilita permeabilizando la célula o tejido, química, electrónica, mecánica o fisiológicamente, mediante lo cual dicho oligonucleótido a ligar puede obtener acceso a un sitio de DSB, y a través de su característica de unión terminal, por ejemplo fosfato, se liga a el sitio de DSB. Preferiblemente, las células se permeabilizan mediante incubación en amortiguador de lisis. Aún más preferiblemente, dicho sitio de DSB es una cola de arginina reparada antes de la ligación con dicho oligonucleótido.

15 Como se apreciará, ligar dicho primer par de oligonucleótidos/adaptador a la DSB antes de un procesamiento adicional garantiza que la DSB identificada sea un evento verdadero y no una consecuencia de las etapas de procesamiento posteriores, y por lo tanto represente un artefacto del procesamiento.

20 En otra realización preferida más de la invención, la parte i) también incluye extraer ADNg de dichas células usando cualquier medio convencional, tal como un amortiguador de extracción o similar, antes de realizar las etapas posteriores.

25 En una realización preferida de la invención, la etapa ii) comprende fragmentar el ADNg en fragmentos más pequeños mediante cualquier medio conocido en la técnica, tal como sonicación o etiquetado.

30 En aún otra realización preferida de la invención, dicho método comprende además una etapa opcional, después de la parte ii) y/o la parte iv), de eliminar fragmentos cuyo tamaño sea menor que alrededor de 100 pb, más preferiblemente menor que alrededor de 150 pb, y retener fragmentos cuyo el tamaño sea mayor que alrededor de 150 pb. Como apreciarán los expertos en la técnica, esta etapa elimina ventajosamente cualesquiera hebras/dímeros oligonucleotídicos que puedan haberse formado y que de otro modo contribuirían posteriormente a artefactos de secuencia. Idealmente, esto se puede realizar usando medios convencionales, tales como el uso de un sonicador Bioruptor y la selección del tamaño usando perlas SPRI (GC Biotech, CNGS-0005) para eliminar fragmentos <150 pb. O seleccione fragmentos en un intervalo preferido tal como, sin limitación, 150-1000 pb, idealmente 200-800 pb, más
35 idealmente 250-750 pb, y aún más preferido 300-500 pb.

40 En aún otra realización preferida de la invención, dicha separación de la parte v) implica usar dicha secuencia de unión proporcionada por el oligonucleótido de la parte i) para unirse a un compañero, y así separar las hebras del Grupo A de la parte iv) de cualesquiera otras hebras. Normalmente, una hebra de unión complementaria a dicha secuencia de unión proporcionada por el oligonucleótido de la parte i) está anclada a un sustrato, y dichas hebras individuales de ácidos nucleicos fluyen por, o sobre, la hebra de unión complementaria anclada.

45 En aún un método preferido adicional de la invención, la parte vi) implica la amplificación del puente, en la que las hebras individuales separadas en la parte v) se amplifican clonalmente sobre un sustrato que tiene oligonucleótidos/sitios de unión anclados en el mismo para la secuencia de unión del primer oligonucleótido de la parte i) y la secuencia adicional del segundo oligonucleótido de la parte iii). De esta manera, se puede unir una única hebra de los fragmentos del Grupo A a ambos extremos de dicho sustrato para facilitar la amplificación del puente. Aún más preferiblemente todavía, dicha secuenciación puede realizarse mediante secuenciación de síntesis empleando el uso de nucleótidos marcados, cada uno de los cuales emite una señal característica que se lee a medida que alarga la
50 secuencia, para proporcionar una lectura de la información de la secuencia. Normalmente, se secuencian varias hebras en un proceso paralelo. Si se prefiere, la secuencia índice puede secuenciarse por separado de la DSB, proporcionando así una indicación de la fuente del ácido nucleico antes de secuenciar la DSB. Alternativamente, los dos índices pueden secuenciarse y de este modo leerse juntos.

55 En una realización particular, el primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido que comprende una secuencia según TCGGTGGTCCGATCATT (SEQ ID NO: 31), y un segundo oligonucleótido que comprende una secuencia según AATGATACGCGACCACCGA (SEQ ID NO: 34).

60 En otra realización, que puede combinarse con la materia objeto del párrafo anterior, el segundo par de oligonucleótidos comprende un primer oligonucleótido de una secuencia que no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30), y un segundo oligonucleótido que comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32).

En una realización particular, el primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido que comprende una secuencia según ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30), y un segundo oligonucleótido que comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32).

5 En otra realización, que puede combinarse con la materia objeto del párrafo anterior, el segundo par de oligonucleótidos puede comprender un primer oligonucleótido que no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31), y un segundo oligonucleótido que comprende una secuencia según AATGATACGCGACCACCGA (SEQ ID NO: 34).

10 SEQ ID NO: 30, 31, 32 y 34 pueden comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

En una realización preferida adicional de la invención, dicho primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido que tiene SEQ ID NO. 1 y un segundo oligonucleótido que tiene SEQ ID NO. 2 o un oligonucleótido que comparte al menos un 80 % de identidad u homología con el mismo, y más preferiblemente, en orden creciente de preferencia, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad u homología con el mismo. Los oligonucleótidos según SEQ ID NO: 1 y SEQ ID NO: 2 pueden comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones. La homología, como se usa aquí, puede denominarse similitud.

20 En una realización preferida adicional de la invención, dicho primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido de la secuencia GATCGGAAGAGCGTCTGTAGGGAAAGAGTGT[ÍNDICE]TCGGTGGGTCGCCGTA TCATTTC, o que comprende de 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2, o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. El primer par de oligonucleótidos de la parte i) puede comprender además un segundo oligonucleótido de la secuencia AATGATACGCGACCACCGA[INDEX]ACACTCTTCCCTACACGACGTCTTCC GATCT, o que comprende 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2, o 1 modificaciones, tales como sustituciones, deleciones, o inserciones.

30 En una realización preferida adicional de la invención, dicho primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido de la secuencia GATCGGAAGAGCGTCTGTAGGGAAAGAGTGT (SEQ ID NO: 37), un índice, y TCGGTGGTCGCCGTATCATTTC (SEQ ID NO: 38). El primer par de oligonucleótidos de la parte i) puede comprender además un segundo oligonucleótido de la secuencia AATGATACGCGACCACCGA (SEQ ID NO: 34), un índice, ACACTCTTCCCTACACGACGCTCTCCGATCT (SEQ ID NO: 39). El índice puede ser cualquier base (n), y puede tener, por ejemplo, una longitud de 5 a 15 o 6 a 10 pares de bases. SEQ ID NO: 34, 37, 38 y 39 pueden comprender de 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

40 En una realización adicional de la invención, dicho primer par de oligonucleótidos de la parte i) comprende un primer oligonucleótido que comprende SEQ ID NO. 3, un índice, y la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30). El primer par de oligonucleótidos de la parte i) puede comprender además un segundo oligonucleótido que comprende, en orden 5' a 3', la secuencia de CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32), un índice, y la secuencia GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 33). El índice puede ser cualquier base (n), y puede tener, por ejemplo, una longitud de 5 a 15 o 6 a 10 pares de bases. SEQ ID NO: 3, 30, 32 y 33 pueden comprender de 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

50 En una realización preferida adicional de la invención, dicho segundo par de oligonucleótidos de la parte iii) comprende un primer oligonucleótido que tiene SEQ ID NO. 3 y un segundo oligonucleótido de secuencia CAAGCAGAAGACGGCATAACGAGAT[INDEX]GTGACTGGAGTTCAGACGTGTGC TCTTCCGATCT o un oligonucleótido que comparte al menos un 80 % de identidad u homología con el mismo, y más preferiblemente, en orden creciente de preferencia, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad u homología con el mismo. El oligonucleótido según SEQ ID NO: 3 puede comprender de 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. El oligonucleótido según CAAGCAGAAGACGGCATAACGAGAT[ÍNDICE]GTGACTGGAGTTCAGACGTGTGC TCTTCCGATCT puede comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, de 1 a 3, 2, o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

60 En una realización preferida adicional de la invención, dicho segundo par de oligonucleótidos de la parte iii) comprende un primer oligonucleótido que tiene SEQ ID NO. 3 y un segundo oligonucleótido que tiene, en orden 5' a 3', la secuencia de CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32), índice, y la secuencia GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 33). El índice puede ser cualquier base (n), y puede tener, por ejemplo, una longitud de 5 a 15 o 6 a 10 pares de bases. Los oligonucleótidos pueden comprender secuencias que tienen 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad u homología con cualquiera de SEQ ID NO: 3, 32 o 33. SEQ ID NO: 3, 32, o 33 pueden comprender de 1 a 7, 1 a 6, 1 a 5, 1 a 4, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

En aún una realización preferida adicional de la invención, dicho segundo oligonucleótido de dicho segundo par de oligonucleótidos de la parte iii) comprende una cualquiera de las siguientes secuencias SEQ ID NO: 4-28, 30-34, 37-39 o 41-42 o un oligonucleótido que comparte al menos un 80 % de identidad u homología con el mismo, y más preferiblemente, en orden creciente de preferencia, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad u homología con el mismo. SEQ ID NO: 4-28, 30-34, 37-39, o 41-42 pueden comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, 1 a 3, 2, o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

En una realización preferida, dicha muestra es una muestra de mamífero, idealmente humana.

Se describe como referencia un kit de piezas, idealmente adecuado para identificar roturas de doble hebra (DSB) de ADN en una muestra de ADNg, que comprende

i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido de este primer par y comprende una característica de unión 3' para la unión a una segunda hebra de dichas DSB; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

ii) un segundo par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, y un sitio de hibridación (RD2 SP) al que se puede unir un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario a dicho primer oligonucleótido de este segundo par y comprende una característica de unión 3' para unirse a una segunda hebra de dichas DSB, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional que es, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

En un kit preferido, dicho kit comprende además al menos un cebador que se une al primer y/o segundo sitio de hibridación con fines de secuenciación.

Aún más preferiblemente, dicho kit comprende además agentes de fragmentación y/o agentes desnaturalizantes para fragmentar y/o desnaturalizar el ácido nucleico en fragmentos y/o hebras individuales, respectivamente.

También se describe como referencia un kit de piezas, adecuado para identificar roturas de doble hebra (DSB) de ADN en una muestra de ADNg, que comprende

i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido de este primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

iii) un segundo par de oligonucleótidos, uno de los cuales es en parte complementario a un segundo oligonucleótido del segundo par; y un segundo oligonucleótido que comprende una característica de unión 3' para la unión a una segunda hebra de dichas DSB, y una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

En un aspecto de la invención, se proporciona un kit adecuado para la preparación de muestras para identificar las DSB en una muestra de ADNg, que comprende

i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según TCGGTGGTCCGATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según

CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

5 En otro aspecto de la invención, se proporciona un kit adecuado para la preparación de muestras para identificar las DSB en una muestra de ADNg, que comprende

10 i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par y opcionalmente comprende una característica de unión 3'; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'; y

15 ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCCGTATCATT (SEQ ID NO: 31), y opcionalmente comprende una característica de unión 5'; y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

SEQ ID NO: 30, 31, 32 y 34 pueden comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En una realización, las modificaciones son sustituciones.

25 El segundo oligonucleótido de la parte i) puede comprender una característica de unión 3', y el primer oligonucleótido de la parte ii) puede comprender una característica de unión 5'. Los primeros oligonucleótidos del primer y segundo par de oligonucleótidos pueden incluir sitios de hibridación (RD1 SP) a los que se pueden unir los cebadores de secuenciación.

30 Se describe como referencia un adaptador de doble hebra, idealmente para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, tal como una muestra de ADNg, que comprende:

35 una primera hebra de oligonucleótido que comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y una segunda hebra de oligonucleótido que es complementaria a dicho primer oligonucleótido y comprende una característica de unión 3' para la unión a una segunda hebra de dichas DSB; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

40 También se describe como referencia un adaptador de doble hebra, idealmente para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, tal como una muestra de ADNg, que comprende:

45 una primera hebra de oligonucleótido que comprende una característica de unión 5', que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, y un sitio de hibridación (RD2 SP) al que se puede unir un cebador de secuenciación; y una segunda hebra de oligonucleótido más larga que es en parte complementario a dicho primer oligonucleótido y comprende una característica de unión 3' para unirse a una segunda hebra de dichas DSB, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional que es, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

50 Según un aspecto adicional de la invención, se proporciona un adaptador de doble hebra, adecuado para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, tal como una muestra de ADNg, que comprende:

55 una primera hebra de oligonucleótidos no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'. El primer oligonucleótido puede comprender una característica de unión 5'.

60 Según un aspecto adicional de la invención, se proporciona un adaptador de doble hebra, adecuado para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, tal como una muestra de ADNg, que comprende:

65

una primera hebra de oligonucleótidos no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34); y

en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'. El primer oligonucleótido puede comprender una característica de unión 5'.

En un aspecto de la invención, se proporciona un método de preparación de muestras para identificar las DSB en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método:

- a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
- b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación;
- c) fragmentar la pluralidad de ácidos nucleicos; y
- d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación. El oligonucleótido del segundo adaptador capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por fragmentación comprende una secuencia idéntica a una región de un segundo cebador.

En algunas realizaciones, los ácidos nucleicos dentro de la muestra que comprende una pluralidad de ácidos nucleicos son bicatenarios durante las etapas b) a d). En otras realizaciones, los ácidos nucleicos pueden ser monocatenarios durante las etapas b) a d). En aún otras realizaciones, los ácidos nucleicos pueden ser bicatenarios para algunas etapas y monocatenarios para otras (por ejemplo, véase la Figura 16). En realizaciones que implican la ligación de un adaptador de doble hebra a un ácido nucleico monocatenario, el adaptador puede comprender un "oligo de alargamiento". El oligo de alargamiento puede comprender nucleótidos aleatorios, tales como 6-8 nucleótidos aleatorios, y colocarse en el extremo 3' del oligonucleótido del adaptador que no se liga a la muestra de ácido nucleico. Los oligos de alargamiento pueden ayudar en el proceso de ligación.

En un aspecto de la invención, se proporciona un método de preparación de muestras para identificar las DSB en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende un primer cebador inmovilizado y para que sean adecuados para la amplificación, que comprende el uso de dicho primer cebador inmovilizado y un segundo cebador, comprendiendo el método:

- a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
- b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación;
- c) fragmentar la pluralidad de ácidos nucleicos; y
- d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que comprende una secuencia idéntica a una región del segundo cebador.

Una muestra de ácido nucleico adecuada para la amplificación que comprende el uso de un primer cebador inmovilizado y un segundo cebador, es una muestra capaz de ser amplificada por un cebador de la misma secuencia que el primer cebador inmovilizado y un cebador de la misma secuencia que el segundo cebador. La idoneidad para la amplificación se puede determinar en disolución.

En algunas realizaciones, el segundo cebador no está inmovilizado al sustrato. Por ejemplo, el sustrato puede ser una perla, y la amplificación puede tener lugar mediante amplificación en emulsión de perlas.

En otras realizaciones, el segundo cebador está inmovilizado en el sustrato. Por ejemplo, el sustrato puede ser una celda de flujo, y la amplificación puede tener lugar mediante amplificación de puente.

En algunas realizaciones, los adaptadores son oligonucleótidos monocatenarios.

En otras realizaciones, los adaptadores comprenden un primer y un segundo oligonucleótido que son al menos parcialmente complementarios. En tales realizaciones, el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, en el que el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de unirse mediante hibridación a un cebador inmovilizado al sustrato. Además, en tales realizaciones, el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y en el que el primer oligonucleótido se puede ligar a un extremo 5' y comprende una secuencia idéntica a una región del segundo cebador, y el segundo oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del segundo cebador.

Por tanto, en una realización, se proporciona un método de preparación de muestras para identificar las DSB de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende un primer cebador inmovilizado y para que sean adecuados para la amplificación, que comprende el uso de dicho primer cebador inmovilizado y un segundo cebador, comprendiendo el método:

- a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
- b) exponer la pluralidad de ácidos nucleicos a un primer par de adaptadores en condiciones propicias para la ligación, en el que el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y en el que el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación.
- c) fragmentar la pluralidad de ácidos nucleicos; y
- d) exponer la pluralidad de ácidos nucleicos a un segundo par de adaptadores en condiciones propicias para la ligación, en el que el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y en el que el primer oligonucleótido se puede ligar a un extremo 5' y comprende una secuencia idéntica a una región del segundo cebador, y el segundo oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del segundo cebador.

En algunas realizaciones, el sustrato comprende un primer cebador inmovilizado y un segundo cebador inmovilizado. Los cebadores inmovilizados pueden, en algunas realizaciones, ser adecuados para actuar como cebadores durante la amplificación del puente.

Por tanto, en una realización, se proporciona un método de preparación de muestras para identificar las DSB de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende un primer y un segundo cebador inmovilizado, comprendiendo el método:

- a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
- b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de unirse al primer cebador inmovilizado mediante hibridación;
- c) fragmentar la pluralidad de ácidos nucleicos; y
- d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que comprende una secuencia idéntica a una región del segundo cebador inmovilizado.

En otras realizaciones, los adaptadores son pares de adaptadores que comprenden un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios. Por tanto, se proporciona un método de preparación de muestras para identificar las DSB de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende un primer y un segundo cebador inmovilizado, comprendiendo el método:

- a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
- b) exponer la pluralidad de ácidos nucleicos a un primer par de adaptadores en condiciones propicias para la ligación, en el que el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y en el que el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, y en el que el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de unirse al primer cebador inmovilizado mediante hibridación;
- c) fragmentar la pluralidad de ácidos nucleicos; y
- d) exponer la pluralidad de ácidos nucleicos a un segundo par de adaptadores en condiciones propicias para la ligación, en el que el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y el primer oligonucleótido que se puede ligar a un extremo 5' comprende una secuencia idéntica a una región del segundo cebador inmovilizado, y en el que el otro oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del segundo cebador inmovilizado.

Como se analiza aquí, en algunas realizaciones, los ácidos nucleicos dentro de la muestra pueden mantenerse como moléculas bicatenarias durante las etapas a) a d). Como tal, en realizaciones particulares, los métodos pueden comprender además:

desnaturalizar la pluralidad de ácidos nucleicos bicatenarios para formar una pluralidad de ácidos nucleicos monocatenarios. Esto puede ser la "etapa e)" en algunas realizaciones.

En una realización particular, los métodos pueden comprender además:

poner en contacto la pluralidad de ácidos nucleicos monocatenarios con el sustrato que comprende cebadores inmovilizados, en condiciones adecuadas para la hibridación de los cebadores inmovilizados con ácidos nucleicos complementarios. Esto puede ser la "etapa f)" en algunas realizaciones.

Las características descritas en relación con la etapa i) de los métodos descritos aquí también son aplicables a la etapa b). Las características descritas en relación con la etapa ii) de los métodos descritos aquí también son aplicables a la etapa c). Las características descritas en relación con la etapa iii) de los métodos descritos aquí también son aplicables a la etapa d). Las características descritas en relación con la etapa iv) de los métodos descritos aquí también son aplicables a la etapa e). Las características descritas en relación con la etapa v) de los métodos descritos aquí también son aplicables a la etapa f).

El sustrato puede ser una superficie sólida, tal como una superficie de una celda de flujo, una perla, un portaobjetos, o una membrana. En particular, el sustrato puede ser una celda de flujo. El sustrato puede ser una celda de flujo con o sin patrón. El sustrato puede comprender vidrio, cuarzo, sílice, metal, cerámica, o plástico. La superficie del sustrato puede comprender una matriz o revestimiento de poli(acrilamida).

Como se usa aquí, se pretende que la expresión "celda de flujo" tenga el significado habitual en la técnica, en particular en el campo de la secuenciación por síntesis. Las celdas de flujo ejemplares incluyen, pero no se limitan a, las usadas en un aparato de secuenciación de ácidos nucleicos, tales como celdas de flujo para las plataformas Genome Analyzer[®], MiSeq[®], NextSeq[®], HiSeq[®], o NovaSeq[®] comercializadas por Illumina, Inc. (San Diego, California); o para la plataforma de secuenciación SOLiD[™] o Ion Torrent[™] comercializada por Life Technologies (Carlsbad, California). También se describen celdas de flujo ejemplares y métodos para su fabricación y uso, por ejemplo, en el documento WO2014/142841 A1; la publicación de solicitud de patente de EE. UU. n.º 2010/0111768 A1, y la patente de EE. UU. n.º 8.951.781.

El sustrato puede comprender cebadores inmovilizados, por ejemplo dos tipos de cebadores que juntos pueden actuar como cebadores directos e inversos para la amplificación del puente. La inmovilización a un sustrato significa que el cebador se une al sustrato incluso en condiciones que desnaturalizarían los ácidos nucleicos bicatenarios. Por ejemplo, el cebador puede estar unido covalentemente al sustrato. Los cebadores están orientados de manera que el extremo 5' sea proximal y el extremo 3' sea distal al punto de inmovilización. Tales disposiciones son estándar en la técnica.

las etapas se pueden realizar en el orden: etapa c), etapa d), y después etapa b). Este orden es particularmente relevante para realizaciones en las que se induce potencialmente una DSB en la muestra, a diferencia de realizaciones para la detección de una DSB preexistente. Con referencia a las declaraciones de la invención que presentan etapas definidas mediante números romanos, las etapas pueden realizarse en el orden: etapa ii), etapa iii), y después etapa i). En tales realizaciones, la DSB puede inducirse después de la ligación del segundo adaptador y antes de la ligación del primer adaptador. Por ejemplo, véase la Figura 13. La inducción de la DSB puede comprender exponer la muestra a condiciones capaces de causar o que se sospecha que pueden causar una DSB.

En realizaciones en las que la etapa d) se lleva a cabo antes de la etapa b), los adaptadores de la etapa d) pueden comprender características de protección de 3' y/o 5' para evitar la ligación de los adaptadores de la etapa b) a los de la etapa d). Las características de protección pueden hacer que el primer adaptador no pueda ligarse al segundo adaptador. En tales realizaciones, el segundo adaptador sigue siendo incapaz de ligarse al primer adaptador debido a que la ligación tiene lugar antes de que el primer adaptador esté presente.

En algunas realizaciones, las etapas se llevan a cabo en el orden: etapa b), etapa c) y etapa d), en el que la etapa b) se lleva a cabo en una célula o *in situ*. En otras realizaciones, las etapas se llevan a cabo en el orden: etapa c), etapa d), y después etapa b), en el que la etapa c) se lleva a cabo *in vitro* después del aislamiento de la muestra de ácido nucleico.

Por lo tanto, en una realización, se proporciona un método de preparación de muestras para identificar las DSB en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método, en el orden:

a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;

c) fragmentar la pluralidad de ácidos nucleicos;

d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación, que comprende opcionalmente una secuencia idéntica a una región del segundo cebador; y

b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB, pero no es capaz de ligarse al segundo adaptador, y que comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación.

En la etapa a), la muestra que comprende una pluralidad de ácidos nucleicos bicatenarios puede ser cualquier muestra de ADN capaz de comprender las DSB, tal como ADNg.

Para cualquiera de los métodos descritos aquí, la muestra puede contener las DSB, o se puede haber tratado de una manera que podría introducir o introduce las DSB.

Por ejemplo, los métodos descritos aquí se pueden usar para detectar daños o cambios en el ADN que se limitan a una sola hebra. Por tanto, los métodos descritos aquí pueden ser para la detección y/o cuantificación de una característica de interés dentro de una muestra de ácido nucleico. Como ejemplo, una lesión en una hebra de ADN bicatenario puede convertirse enzimáticamente en una DSB, que entonces puede detectarse mediante los métodos descritos aquí.

En algunos ejemplos, la lesión puede ser una rotura de una sola hebra. En otros ejemplos, la lesión es un cambio de base en una muestra de ácido nucleico. Por ejemplo, los métodos descritos aquí pueden ser para la detección de la edición de bases inducida por CRISPR/Cas, tal como un editor de bases de citosina o un editor de bases de adenosina. Dichas ediciones pueden convertirse en DSB y detectarse como se describe aquí. La Figura 17 y la Figura 18 ilustran realizaciones ejemplares.

Por tanto, en una realización, el método comprende proporcionar una muestra que comprende una pluralidad de ácidos nucleicos que comprenden o se sospecha que comprenden una característica de interés o una lesión; y exponer la muestra a condiciones capaces de convertir la característica de interés o lesión en una DSB.

Para tales realizaciones, el método puede estar en el orden: conversión de la lesión en una DSB, b), c) y d); o puede ser en el orden: c), d), conversión de la lesión en una DSB, y b).

En otras realizaciones, la muestra puede haber sido tratada con una nucleasa, tal como una nucleasa efectora similar a un activador de la transcripción (TALEN), una endonucleasa CRISPR/Cas, una nucleasa de dedos de zinc, una meganucleasa, o cualquier endonucleasa de restricción, y los métodos pueden ser para la detección de efectos fuera de la diana. La muestra puede tratarse con un agente, tal como un agente terapéutico potencial, para determinar si dicho agente es capaz de causar las DSB o las DSB fuera de la diana.

Los métodos descritos aquí pueden usarse para identificar sitios de proteínas que se unen a un ácido nucleico. Por ejemplo, una muestra se puede poner en contacto con un agente de unión a proteínas, tal como un anticuerpo, específico para una proteína de interés, en el que la proteína de interés está potencialmente unida al ADN. El ADN puede ser una muestra que se ha puesto en contacto con la proteína de interés. El agente de unión a proteínas puede estar asociado directa o indirectamente con una nucleasa, formando así una DSB en cualquier sitio al que se une la proteína de interés. Cualesquiera DSB pueden entonces detectarse mediante los métodos descritos aquí. La metodología puede ser los métodos de Escisión Bajo Dianas y Liberación Usando Nucleasas (CUT&RUN). Como tal, el método puede comprender: etapa a); poner en contacto la muestra con una proteína de interés; poner en contacto la muestra con una nucleasa capaz de asociarse directa o indirectamente con la proteína de interés para formar una DSB en los ácidos nucleicos a los que está unida la proteína; etapa b), etapa c), y después etapa d). Este orden es particularmente útil para realizaciones en las que el método hasta la etapa b) se lleva a cabo en una célula. Alternativamente, el método puede comprender: etapa a); etapa c); etapa d); poner en contacto la muestra con una proteína de interés; poner en contacto la muestra con una nucleasa capaz de asociarse directa o indirectamente con la proteína de interés para formar una DSB en los ácidos nucleicos a los que está unida la proteína; y después etapa b). Este orden es particularmente útil para realizaciones *in vitro*.

En otras realizaciones, la DSB puede inducirse deliberadamente en un sitio conocido o diana. Por ejemplo, se puede usar una nucleasa dirigida, tal como CRISPR/Cas o TALEN, para inducir una DSB en una secuencia de interés de modo que los métodos de la invención se puedan usar para aislar dicha secuencia de interés para análisis adicionales. La secuencia de interés puede ser genes específicos, el exoma completo, o un locus específico en el genoma. Como tal, en una realización, los métodos comprenden poner en contacto la muestra con una nucleasa capaz de la inducción dirigida de una DSB en una secuencia de interés.

Los métodos de la invención pueden ser para la detección de eventos de inserción viral o bacteriana, o daños en el ADN causados por eventos de inserción viral o bacteriana. Estos eventos se pueden medir en virtud de las secuencias genéticas únicas asociadas con bacterias y virus que se insertarían en el genoma. Por lo tanto, los métodos de la invención pueden revelar estos sitios de inserción de ADN extraño, que pueden ocurrir a través de una estructura intermedia de DSB.

En otra realización, los métodos descritos aquí se pueden usar para evaluar los riesgos asociados con un agente terapéutico, tal como una terapia génica. Por ejemplo, los métodos descritos aquí se pueden aplicar a una muestra tomada de un paciente, en los que la muestra se ha tratado con el agente terapéutico, para determinar el riesgo, la naturaleza o la frecuencia de los efectos fuera de diana para dicho paciente. Por lo tanto, los métodos descritos aquí pueden ser útiles para la medicina personalizada al proporcionar un patrón de DSB específico del paciente, y su frecuencia, inducido por un agente de interés. Como tal, en una realización, los métodos pueden comprender obtener una muestra de un sujeto y exponer dicha muestra a un agente, tal como un agente terapéutico. El método puede comprender determinar la naturaleza y/o frecuencia de cualquier lesión o DSB en la muestra después de la exposición. El orden de las etapas de la invención puede ser cualquiera de los descritos aquí.

Los métodos descritos aquí se pueden usar para detectar la contaminación de una muestra por cualquier agente capaz de causar las DSB.

Los métodos descritos aquí se pueden usar para medir la estabilidad de genomas sintéticos o ensamblados artificialmente.

Los métodos descritos aquí se pueden usar para la Evaluación de Riesgos de Próxima Generación (NGRA) en toxicología genética. NGRA se define como un enfoque de evaluación de riesgos basado en hipótesis y basado en exposición que integra nuevas metodologías de enfoque (NAM) para garantizar la seguridad sin el uso de pruebas con animales. Las DSB son una medida directa de la exposición genotóxica, y pueden cuantificarse mediante los métodos de la invención. Por lo tanto, los métodos de la invención pueden usarse para evaluaciones de riesgos que requieran la cuantificación de la exposición genotóxica.

En resumen, la muestra de ácido nucleico que se sospecha que contiene las DSB puede contener dichas DSB debido a un daño natural en el ADN, debido al tratamiento con un agente potencialmente causante de DSB, debido a la inducción deliberada de una DSB en un sitio de interés, o por cualquier otra razón.

En la etapa b), las condiciones permiten la ligación del primer adaptador a una DSB. El primer adaptador comprende un oligonucleótido capaz de ligarse al extremo 3' de una hebra de una DSB, y también puede comprender otro oligonucleótido capaz de ligarse al extremo 5' de una hebra de una DSB. Como tal, el primer adaptador está unido covalentemente a la DSB. La ligación puede ser directa o indirecta. Por ejemplo, la ligación puede realizarse con

nucleótidos adicionales introducidos en la DSB. Alternativamente, un oligonucleótido o un par de oligonucleótidos se puede ligar a la DSB, y el primer adaptador se puede ligar a dicho oligonucleótido u oligonucleótidos. Los métodos de ligación pueden ser cualquiera de los descritos aquí. Los dos oligonucleótidos del primer par de adaptadores pueden ser completamente complementarios.

5 El primer y/o segundo oligonucleótidos del primer par de adaptadores pueden comprender una característica de protección de 3' y/o 5'. Estas características de protección pueden ser cualquiera de las descritas aquí, particularmente cualquiera descrita con respecto al primer par de oligonucleótidos discutidos en relación con la etapa i) de los métodos descritos aquí.

10 El primer y/o segundo oligonucleótidos del primer par de adaptadores pueden comprender cualquier característica descrita con respecto al primer par de oligonucleótidos discutidos en relación con la etapa i) de los métodos descritos aquí. En particular, el primer par de adaptadores puede incluir o no las secuencias descritas con respecto al primer par de oligonucleótidos discutidos en relación con la etapa i).

15 El primer adaptador incluye un oligonucleótido que se puede ligar a un extremo 3', y comprende una secuencia que es capaz de unirse a un cebador inmovilizado mediante hibridación. Esto significa que, cuando un primer par de adaptadores ligado se desnaturaliza en hebras individuales, una hebra es complementaria a un cebador inmovilizado en el sustrato. El adaptador incluye una longitud suficiente de secuencia complementaria para permitir la unión que no se libera durante las etapas de lavado o polimerización. La longitud de la región complementaria puede ser de 5, 10, 15, 20, 21, 24, o más bases. Alternativamente, la región complementaria puede incluir 5, 10, 15, 20, 21, 24, o más bases complementarias.

25 La etapa c) puede comprender cualquier método de fragmentación como se analiza aquí.

En la etapa d), las condiciones permiten la ligación de los ácidos nucleicos a un segundo adaptador. El segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' en un sitio de fragmentación, y puede comprender un oligonucleótido capaz de ligarse a un extremo 3' en un sitio de fragmentación. Los métodos de ligación pueden ser cualquiera de los descritos aquí.

30 El segundo adaptador no es capaz de ligarse al primer adaptador. En una realización particular, esta prevención se debe a la inclusión de una característica de protección de 3' en el primer adaptador. En realizaciones relevantes, esta prevención se debe a la inclusión de características de protección de 5' y/o 3' en el primer par de adaptadores. En particular, el primer oligonucleótido del primer adaptador (es decir, el oligonucleótido capaz de unirse a un extremo 3') puede comprender una característica de protección de 3', tal como un terminador de cadena 3' espaciador C3, para evitar la ligación del adaptador a esta hebra (véase la Figura 12).

35 El primer y/o segundo oligonucleótidos del segundo par de adaptadores pueden comprender una característica de protección de 3' y/o 5'. Estas características de protección pueden ser cualquiera de las descritas aquí, particularmente cualquiera descrita con respecto al segundo par de oligonucleótidos discutidos en relación con la etapa iii) de los métodos descritos aquí.

45 En realizaciones en las que la etapa d) se lleva a cabo antes de la etapa b), el primer y/o segundo oligonucleótidos del segundo par de adaptadores pueden comprender una característica de protección de 3' y/o 5', y al menos una característica de protección puede evitar la ligación de los adaptadores de la etapa b) a los de la etapa d). Por ejemplo, el oligonucleótido del segundo adaptador capaz de unirse a un extremo 3' puede comprender una característica de protección de 3', tal como un terminador de cadena 3' espaciador C3.

50 El primer y/o segundo oligonucleótidos del segundo par de adaptadores pueden comprender cualquier característica descrita con respecto al segundo par de oligonucleótidos discutidos en relación con la etapa iii) de los métodos descritos aquí. En particular, el segundo par de adaptadores puede incluir o no las secuencias descritas con respecto al segundo par de oligonucleótidos discutidos en relación con la etapa iii).

55 El segundo par de adaptadores puede introducirse mediante etiquetado. Como tal, la fragmentación y ligación del segundo adaptador puede ser la misma etapa. Por ejemplo, se pueden combinar las etapas c) y d).

60 El segundo adaptador incluye un oligonucleótido que se puede ligar a un extremo 5', y que comprende una secuencia idéntica a al menos parte de un cebador inmovilizado en el sustrato. Por tanto, cuando este oligonucleótido actúa como molde durante la polimerización, la nueva hebra incluirá una secuencia que es capaz de hibridarse con dicho cebador. Por tanto, la longitud de la región relevante de la secuencia adaptadora y la región relevante del cebador deberían ser adecuadas para esta función. La longitud de la región idéntica puede ser de 5, 10, 15, 20, 21, 24, o más bases. El otro oligonucleótido dentro del segundo adaptador no incluye una secuencia complementaria a esta denominada región idéntica. Como tal, el segundo adaptador no es capaz de unirse al sustrato mediante hibridación.

65 Después de la etapa f), los métodos pueden comprender además poner en contacto cualquier ácido nucleico hibridado con una polimerasa en condiciones adecuadas para el alargamiento del cebador inmovilizado para sintetizar un ácido

nucleico que es una cadena de nucleótidos que son complementarios al ácido nucleico hibridado. Después, el ácido nucleico recién formado se puede amplificar. En algunas realizaciones, el cebador para la amplificación también está inmovilizado en el sustrato, y por ejemplo puede ser adecuado para la amplificación del puente. Este proceso es conocido en la técnica, y forma grupos clonales de ácidos nucleicos. En otras realizaciones, el cebador para amplificación puede estar en disolución, por ejemplo para algunas realizaciones en las que el sustrato es una perla. Los ácidos nucleicos amplificados pueden secuenciarse entonces de la manera habitual, por ejemplo mediante secuenciación por síntesis. Los adaptadores pueden comprender un sitio para la unión de un cebador de secuenciación para ayudar en este proceso. Los adaptadores también pueden comprender un índice como se describe aquí.

Por tanto, en una realización, los métodos pueden comprender además:

g) obtener información de secuencia para cualquier ácido nucleico que se hibridó con el sustrato en la etapa f).

Los métodos que incluyen la etapa g) pueden denominarse métodos para identificar las DSB de ADN en una muestra de ácido nucleico.

Las características descritas en relación con la etapa vi) de los métodos descritos aquí también son aplicables al etapa g).

En cualquier realización de productos o métodos que comprenden un oligonucleótido que comprende AATGATACGCGACCACCGA (SEQ ID NO: 34), o variantes del mismo, el oligonucleótido puede comprender AATGATACGCGACCACCGAGATCTACAC (SEQ ID NO: 41), o variantes como se define para SEQ ID NO: 34. En cualquier realización de productos o métodos que comprenden un oligonucleótido que comprende SEQ ID NO: 31 o que comprende no más de 5, 10, 15, o 20 bases de la secuencia TCGGTGGTGCCTGATCATT (SEQ ID NO: 31), el oligonucleótido puede comprender SEQ ID NO: 38 o 42, o puede comprender no más de 5, 10, 15, 20, 25, o 29 bases de la secuencia GTGTAGATCTCGGTGGTGCCTGATCATT (SEQ ID NO: 42), o no más de 5, 10, 15, o 19 bases de la secuencia TCGGTGGTGCCTGATCATT (SEQ ID NO: 38).

La Figura 16 describe un flujo de trabajo en el que se identifica una característica de interés, tal como una lesión en una única hebra de ADN bicatenario. La característica de interés puede ser cualquier característica capaz de escindirse específicamente. Por ejemplo, cualquier característica que dé como resultado la escisión de una sola hebra de ADN bicatenario en el sitio de la característica de interés. La característica de interés puede ser, por ejemplo, un dímero de ciclobutano-pirimidina (CPD), 8-oxoguanina, un sitio abásico, y cualquier combinación de los mismos. La hebra de ADN que comprende la característica de interés puede escindirse, y la muestra bicatenaria desnaturalizarse, para dar como resultado un extremo 3' al que puede ligarse un adaptador.

Por lo tanto, descrito como referencia, se proporciona un método de preparación de muestras para identificar una característica de interés en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados con una característica de interés para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método:

a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos, exponer la pluralidad de ácidos nucleicos a condiciones capaces de escindir al menos una hebra de un ácido nucleico en una característica de interés, y desnaturalizar la pluralidad de ácidos nucleicos en ácidos nucleicos monocatenarios;

b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de un sitio de escisión y que comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación;

c) fragmentar la pluralidad de ácidos nucleicos; y

d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación. El oligonucleótido del segundo adaptador capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por fragmentación puede comprender una secuencia idéntica a una región de un segundo cebador.

Todas las características descritas en relación con los métodos de identificación de las DSB de ADN también son relevantes para los métodos para identificar una característica de interés. En particular, las etapas b), c) y d) del método de preparación de muestras para identificar una característica de interés en una muestra de ácido nucleico pueden ser las mismas que las etapas b), c) y d) como se describe para el método de preparación de muestras para identificar las DSB de ADN en una muestra de ácido nucleico. Los adaptadores pueden ser los mismos que los descritos para identificar las DSB, y los métodos pueden ser para preparar una biblioteca adecuada para hibridarse con cualquier

sustrato descrito aquí. Para realizaciones que comprenden la ligación de un adaptador de doble hebra a un ácido nucleico monocatenario, se puede incluir un oligo de alargamiento, como se describe aquí.

Las etapas se pueden realizar en el orden: c), d), a), y después b). Este es el orden que se muestra en la Figura 16. Por tanto, se proporciona como referencia un método de preparación de muestras para identificar una característica de interés en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados con una característica de interés para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método, en el orden:

c) fragmentar una muestra que comprende una pluralidad de ácidos nucleicos; y

d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación, que comprende opcionalmente una secuencia idéntica a una región del segundo cebador;

a) exponer la pluralidad de ácidos nucleicos a condiciones capaces de escindir al menos una hebra de un ácido nucleico en una característica de interés, y desnaturalizar la pluralidad de ácidos nucleicos en ácidos nucleicos monocatenarios;

b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de un sitio de escisión, pero no es capaz de ligarse al segundo adaptador, y que comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación.

Todas las etapas y características posteriores descritas en relación con los métodos de identificación de las DSB de ADN también son relevantes para los métodos para identificar una característica de interés. En particular, el método puede comprender desnaturalizar la muestra en ácidos nucleicos monocatenarios, por ejemplo para desnaturalizar cualesquiera adaptadores de doble hebra. El método puede comprender además poner en contacto la pluralidad de ácidos nucleicos con el sustrato que comprende cebadores inmovilizados, en condiciones adecuadas para la hibridación de los cebadores inmovilizados con ácidos nucleicos complementarios. Además, el método puede comprender además obtener información de secuencia para cualesquiera ácidos nucleicos hibridados con el sustrato.

En otro ejemplo, se proporciona como referencia un método de preparación de muestras para identificar una característica de interés en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados con una característica de interés para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método:

α) exponer una muestra que comprende una pluralidad de ácidos nucleicos a condiciones capaces de escindir al menos una hebra de un ácido nucleico en una característica de interés, y desnaturalizar la pluralidad de ácidos nucleicos en ácidos nucleicos monocatenarios;

β) exponer la muestra, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5';

γ) fragmentar el ácido nucleico de dicha muestra en fragmentos; y

δ) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende un sitio de hibridación (RD2 SP) al que se puede unir un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario de dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para unirse a una segunda hebra de dicho ácido nucleico fragmentado, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional que es, opcionalmente, una secuencia de unión para permitir la amplificación del puente; y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'.

El orden de las etapas puede ser etapa γ), etapa δ, etapa α), y después etapa β). Así, en otro ejemplo de referencia, se proporciona un método de preparación de muestras para identificar una característica de interés en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados con una característica de interés para que sean adecuados para unirse a un sustrato que comprende cebadores inmovilizados, comprendiendo el método, en el orden:

y) fragmentar una muestra de ácido nucleico en fragmentos de ácido nucleico;

5 δ) exponer dichos fragmentos, en condiciones de ligación, a un segundo par de oligonucleótidos, el primero de los cuales comprende un sitio de hibridación (RD2 SP) al que se puede unir un segundo cebador de secuenciación; y un segundo oligonucleótido más largo que es en parte complementario de dicho primer oligonucleótido del segundo par y comprende una característica de unión 3' para unirse a una segunda hebra de dicho ácido nucleico fragmentado, una secuencia complementaria a dicho sitio de hibridación, y una secuencia adicional que es, opcionalmente, una secuencia de unión para permitir la amplificación del puente;

10 y en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'.

15 α) exponer la muestra a condiciones capaces de escindir al menos una hebra de un ácido nucleico en una característica de interés, y desnaturalizar la pluralidad de ácidos nucleicos en ácidos nucleicos monocatenarios; y

20 β) exponer la muestra, en condiciones de ligación, a un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la ligación de dicho oligonucleótido a una primera hebra de dicha DSB, un sitio de hibridación (RD1 SP) al que se puede unir un primer cebador de secuenciación y una secuencia de unión para separar dicha DSB de un conjunto de DSB; y un segundo oligonucleótido que es complementario a dicho primer oligonucleótido del primer par; en el que uno o ambos de dichos oligonucleótidos comprenden una característica de protección de 3' y/o 5'.

25 El método puede comprender además:

25 e) desnaturalizar los fragmentos para proporcionar ácidos nucleicos monocatenarios.

El método puede comprender además:

30 ζ) separar las hebras de la parte ε) en dos grupos: grupo A: aquellos fragmentos que tienen ligado en un primer extremo el primer sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte β), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte δ); y grupo B: aquellos fragmentos que no tienen ligados en un primer extremo el sitio de hibridación y la secuencia de unión proporcionada por el oligonucleótido de la parte β), y en otro extremo, el segundo sitio de hibridación y la secuencia adicional proporcionada por el oligonucleótido de la parte δ).

35

El método puede comprender además:

40 η) secuenciar las hebras del grupo A usando cebadores que se unen al primer y/o segundo sitio de hidridación.

Los métodos pueden adaptarse para que sean particularmente adecuados para uso con sistemas basados en perlas. Por ejemplo, la secuenciación de Ion Torrent. Un ejemplo de referencia particular se ejemplifica en las figuras 14 y 15.

45 Como tal, se proporciona como referencia un método de preparación de muestras para identificar las DSB de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean adecuados para unirse a un sustrato que comprende un primer cebador inmovilizado, comprendiendo el método:

50 1) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;

2) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de hibridarse a un segundo cebador;

55 3) fragmentar la pluralidad de ácidos nucleicos;

4) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que comprende una secuencia idéntica a una región del primer cebador inmovilizado; y

60

5) poner en contacto la pluralidad de ácidos nucleicos con el segundo cebador en condiciones adecuadas para el alargamiento del cebador.

65 La muestra que comprende una pluralidad de ácidos nucleicos puede ser cualquiera de las descritas aquí. En particular, se puede sospechar que la muestra contiene las DSB, se puede tratar con agentes capaces de causar o

que se sospecha que pueden causar las DSB, o puede comprender o sospecharse que comprende una característica de interés que se puede convertir en una DSB. La muestra puede incluir una característica de interés capaz de escindirse específicamente.

5 La muestra puede ser cualquier muestra de ADN capaz de comprender las DSB, por ejemplo ADNg.

En un ejemplo de referencia particular, la etapa 2) es:

10 exponer la pluralidad de ácidos nucleicos a un primer par de adaptadores en condiciones propicias para la ligación, en el que el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y en el que el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de hibridarse con un segundo cebador; y

15 en el que la etapa 4) es:

20 exponer la pluralidad de ácidos nucleicos a un segundo par de adaptadores en condiciones propicias para la ligación, en el que el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 5' y comprende una secuencia idéntica a una región del primer cebador inmovilizado, y el segundo oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del primer cebador inmovilizado.

25 En ejemplos particulares, el oligonucleótido del segundo par de adaptadores que se puede ligar a un extremo 5' comprende 5, 10, 15, 20, o las 23 bases de la secuencia según AACCCACTACCGCCTCCGCTTTCC (SEQ ID NO: 40). El otro oligonucleótido es de una secuencia que no comprende una secuencia de más de 5, 10, 15, 20 bases, o no comprende las 22 bases, de la secuencia GGAAAGCGGAGGGCGTAGTGGTT (SEQ ID NO: 36).

30 Los oligonucleótidos según SEQ ID NO: 36 y SEQ ID NO: 40 pueden comprender de 1 a 12, 1 a 10, 1 a 8, 1 a 5, 1 a 3, 2 o 1 modificaciones, tales como sustituciones, deleciones, o inserciones. En un ejemplo, las modificaciones son sustituciones.

35 El primer y/o segundo oligonucleótidos del primer par de adaptadores pueden comprender una característica de protección de 3' y/o 5'. Estas características de protección pueden ser cualquiera de las descritas aquí, particularmente cualquiera descrita con respecto al primer par de oligonucleótidos discutidos en relación con la etapa i) de los métodos descritos aquí.

40 El primer y/o segundo oligonucleótidos del segundo par de adaptadores pueden comprender una característica de protección de 3' y/o 5'. Estas características de protección pueden ser cualquiera de las descritas aquí, particularmente cualquiera descrita con respecto al segundo par de oligonucleótidos discutidos en relación con la etapa iii) de los métodos descritos aquí.

45 En ejemplos particulares, el segundo adaptador no puede ligarse al primer adaptador debido a la presencia de una modificación 3' del primer adaptador. Por ejemplo, un terminador de cadena 3' C3.

50 El oligonucleótido del segundo adaptador que se puede ligar a un extremo 5' comprende una secuencia idéntica a una región de un segundo cebador de modo que, cuando se genera una hebra complementaria, la hebra complementaria comprende una región a la que el cebador puede unirse mediante hibridación. Esta secuencia idéntica a una región de un segundo cebador puede ser idéntica a 5, 10, 15, 20, 21, 24, o más bases de un cebador inmovilizado.

55 El método puede comprender además desnaturalizar la pluralidad de ácidos nucleicos para formar una pluralidad de ácidos nucleicos monocatenarios. Las características de este etapa pueden ser cualquiera de las descritas aquí en relación con las realizaciones de la invención.

60 El método puede comprender además poner en contacto la pluralidad de ácidos nucleicos con el sustrato que comprende el primer cebador inmovilizado en condiciones adecuadas para la hibridación del primer cebador inmovilizado con ácidos nucleicos complementarios. Por ejemplo, la muestra de ácidos nucleicos con adaptadores ligados puede unirse entonces al sustrato, tal como por ejemplo una perla que comprende cebadores inmovilizados. Los cebadores pueden inmovilizarse en la perla de manera que el extremo 5' sea proximal y el extremo 3' sea distal al punto de inmovilización.

65 Se pueden usar técnicas habituales de manera que el sustrato, tal como la perla, presente múltiples copias de un ácido nucleico con la misma secuencia.

Los métodos pueden comprender además obtener información de secuencia para cualesquiera ácidos nucleicos hibridados con el sustrato. Los adaptadores pueden comprender sitios para la unión de cebadores de secuenciación, para ayudar con este proceso. Los adaptadores pueden comprender secuencias de índice para ayudar con este proceso.

5 Se proporciona como referencia un adaptador de doble hebra, adecuado para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, tal como una muestra de ADNg, que comprende:

10 un primer oligonucleótido que comprende una secuencia según AACCCACTACGCCCTCCGCTTTCC (SEQ ID NO: 40); y

15 un segundo oligonucleótido de una secuencia que no comprende una secuencia de más de 5, 10, 15, 20 bases, o no comprende las 22 bases, de la secuencia GGAAAGCGGAGGGCGTAGTGGTT (SEQ ID NO: 36); en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

20 En las reivindicaciones que siguen y en la descripción anterior de la invención, excepto cuando el contexto requiera lo contrario debido al lenguaje expreso o implicación necesaria, la palabra "comprende", o variaciones tales como "comprende" o "que comprende", se usa en un sentido inclusivo, es decir, para especificar la presencia de las características indicadas pero no para excluir la presencia o adición de características adicionales en diversas realizaciones de la invención.

25 Las características preferidas de cada aspecto de la invención pueden ser las descritas en relación con cualquiera de los otros aspectos.

30 Cualquier característica descrita en relación con las declaraciones de invención que presentan las etapas i), ii), iii), etc., puede combinarse con características descritas en relación con las declaraciones de invención que presentan las etapas a), b), c), etc. Cualesquiera características descritas en relación con las declaraciones de invención que presentan las etapas a), b), c), etc. pueden combinarse con características descritas en relación con las declaraciones de invención que presentan las etapas i), ii), iii), etc.

35 Otras características de la presente invención resultarán evidentes a partir de los siguientes ejemplos. La presente invención se describirá ahora a modo de ejemplo únicamente, con referencia particular a las siguientes figuras, en las que:

EJEMPLOS

MÉTODOS

40 Cultivo y tratamiento celular

45 Se cultivaron células HEK293, HEK293T y U2OS DivA en DMEM (Life Technologies) suplementado con 10% de FBS (Life Technologies) a 37 °C y 5 % de CO₂. Las células HEK293 se nucleofectaron con 224 pmol de Ribonucleoproteína (RNP) por 3,5 x 10⁵ células usando una unidad Lonza 4D-Nucleofector X con código de pulso CM-130. Las células se recolectaron a 0, 7, 12, 24 y 30 h después de la nucleofección para el procesamiento INDUCE-seq. Para estimular la inducción de las DSB dependiente de AsiSI, las células DivA se trataron con 4OHT 300 nM (Sigma, H7904) durante 4 h.

50 Proteína Cas9 y sgRNA

El ARN guía dirigido contra EMX1 (GAGTCCGAGCAGAAGAAGAA; SEQ ID NO: 29) se sintetizó como un oligonucleótido de sgRNA no modificado de longitud completa (Synthego). La proteína Cas9 se produjo en nuestras instalaciones (AstraZeneca), y contenía una etiqueta 6xHN N-terminal.

55 Método INDUCE-seq

60 Las células se sembraron en placas de 96 pocillos previamente revestidas con poli-D-lisina (Greiner bio-one, 655940) a una densidad de ~1 x 10⁵/pocillo, y se reticularon en paraformaldehído (PFA) al 4 % (Pierce, 28908) durante 10 min a temperatura ambiente (ta). Las células se lavaron en 1x PBS para eliminar el formaldehído, y se almacenaron a 4 °C durante hasta 30 días. El método INDUCE-seq se inició permeabilizando células. Entre las etapas de incubación, las células se lavaron en 1x PBS a ta. Las células se permeabilizaron mediante incubación en amortiguador de lisis 1 (Tris-HCl 10 mM, pH 8, NaCl 10 mM, EDTA 1 mM, Triton X-100 al 0,2 %, pH 8 a 4 °C) durante una hora a ta, seguido de incubación en amortiguador de lisis 2 (Tris-HCl 10 mM, NaCl 150 mM, EDTA 1 mM, SDS al 0,3 %, pH 8 a 25 °C) durante una hora a 37 °C. Las células permeabilizadas se lavaron tres veces en amortiguador CutSmart® 1x (NEB, B7204S), y los extremos romos se repararon usando el kit NEB Quick Blunting Kit (E1201L) + 100 µg/ml de BSA en un volumen final de 50 µl a ta durante una hora. Las células se lavaron entonces tres veces en amortiguador CutSmart®

1x, y se colocaron colas A para añadir un único dATP al extremo 3' del ADN bicatenario usando el módulo NEBNext® dA-Tailing Module (NEB, E6053L) en un volumen final de 50 µl a 37 °C durante 30 minutos. Las células con colas A se lavaron tres veces en amortiguador CutSmart® 1x, y después se incubaron en amortiguador de ADN ligasa T4 1x (NEB, B0202S) durante 5 minutos a ta. Los extremos con colas A se marcaron mediante ligación usando la ADN ligasa T4 (NEB, M0202M) + 0,4 µM de adaptador P5 en un volumen final de 50 µl a 16 °C durante 16-20 h. Después de la ligación, el exceso de adaptador P5 se eliminó lavando las células 10 veces en amortiguador de lavado a ta (Tris-HCl 10 mM, NaCl 2 M, EDTA 2 mM, Triton X-100 al 0,5 %, pH 8 a 25 °C), incubando durante 2 minutos en cada etapa de lavado. Las células se lavaron una vez en PBS, y después una vez en H₂O libre de nucleasa (IDT, 05-11-01-04). El ADN genómico se extrajo incubando células en amortiguador de extracción de ADN (Tris-HCl 10 mM, NaCl 100 mM, EDTA 50 mM, SDS al 1,0 %, pH 8 a 25 °C) + Proteinasa K 1 mg/ml (Invitrogen, AM2584) en un volumen final de 100 µl durante 5 minutos a ta. Los lisados celulares se transfirieron a tubos Eppendorf RNA/DNA LoBind de 1,5 ml (Fisher Scientific, 13-698-792), y se incubaron a 65 °C durante 1 hora, agitando a 800 rpm. El ADN se purificó usando Genomic DNA Clean & Concentrator™-10 (Zymo Research, D4010), y se eluyó usando 100 µl de amortiguador de elución. El rendimiento de ADN se evaluó usando 1 µl de muestra y el kit Qubit DNA HS (Invitrogen, Q32854) antes de continuar con la preparación de la biblioteca. El ADN genómico se fragmentó hasta 300-500 pb usando un sonicador Bioruptor, y el tamaño se seleccionó usando perlas SPRI (GC Biotech, CNGS-0005) para eliminar fragmentos <150 pb. El ADN fragmentado y de tamaño seleccionado se reparó en sus extremos usando el módulo de ADN NEBNext® Ultra™ II (NEB, E7546L). El ADN fragmentado y reparado en los extremos se añadió directamente a la reacción de ligación usando el módulo de ligación NEBNext® Ultra™ II (NEB, E7595L) según las instrucciones del fabricante, usando 7,5 µM de adaptador P7 semifuncional modificado y omitiendo la adición de enzima USER. Las bibliotecas de secuenciación ligadas se purificaron usando perlas SPRI. Las bibliotecas se purificaron dos veces más usando perlas SPRI, y el tamaño se seleccionó para eliminar fragmentos <200 pb, para eliminar el ADN adaptador residual. Las bibliotecas limpias finales se cuantificaron mediante qPCR usando el kit de cuantificación de bibliotecas KAPA para plataformas Illumina® (Roche, 07960255001). Las muestras se combinaron y concentraron hasta el volumen deseado para secuenciación usando un SpeedVac. La secuenciación se realizó en un Illumina NextSeq 550 usando una celda de flujo de alta capacidad de 1x75 pb.

Adaptadores INDUCE-seq

30 Todos los oligonucleótidos adaptadores INDUCE-seq modificados se adquirieron de IDT. Los oligonucleótidos monocatenarios se hibridaron a una concentración final de 10 µM en amortiguador dúplex sin nucleasas (IDT, 11-01-03-01) calentando a 95 °C durante 5 minutos y enfriando lentamente hasta 25 °C usando un termociclador. En la figura 2 se proporciona una descripción general de la estructura de los oligonucleótidos adaptadores.

35 Inducción de DSB *in situ* con HindIII

Los experimentos piloto INDUCE-seq se llevaron a cabo induciendo las DSB *in situ* en células HEK293T usando la enzima de restricción HindIII-HF® (NEB, R3104S). Este proceso fue el mismo que el descrito para el método INDUCE-seq completo, con la adición de inducción de DSB antes del achatamiento final. Después de la permeabilización celular, las DSB se indujeron usando 50 U de HindIII-HF® en amortiguador CutSmart® 1x en un volumen final de 50 µl. Las digestiones se llevaron a cabo a 37°C durante 18 horas.

Esquema de análisis de datos INDUCE-seq

45 Se obtuvieron archivos FASTQ demultiplexados, y se pasaron a través de Trim Galore! (Krueger 2015) para eliminar la secuencia del adaptador en el extremo 3' de las lecturas usando las configuraciones predeterminadas. La calidad de los datos de secuenciación se evaluó mediante FastQC (Andrews 2010). Después de leer el alineamiento con el genoma humano de referencia (GRCh37/hg19) usando BWA-mem (Li y Durbin 2009), los alineamientos cartografiados con una puntuación de alineación baja (MAPQ<30) se eliminaron usando SAMtools (Li *et al.* 2009), y las lecturas poco recortadas se filtraron usando un script AWK personalizado para garantizar una asignación de DSB precisa. Los archivos BAM resultantes se convirtieron en archivos BED usando la función bam2bed de bedtools (Quinlan y Hall 2010), después de lo cual la lista de coordenadas leídas se filtró usando regiones de mala capacidad de cartografiado, extremos de cromosomas, y contiguos del genoma de referencia incompletos, para eliminar estas características de los datos. Las posiciones de las DSB se asignaron como el primer nucleótido 5' en dirección 5' de la lectura con respecto a la orientación de la hebra, y se generaron como un archivo BED 'breakends'. Se tuvo cuidado de eliminar duplicados ópticos conservando al mismo tiempo los eventos de DSB recurrentes reales. Al mantener cada ID de lectura, se usó la información posicional de las celdas de flujo X e Y para filtrar duplicados ópticos mediante un script AWK personalizado. El resultado final fue un archivo BED que contenía una lista de posiciones de rotura de nucleótidos individuales cuantificadas.

60 Análisis de las DSB inducidas por HindIII en células HEK293T

Las posiciones de los sitios diana de HindIII dentro de hg19 se predijeron primero *in silico* usando la herramienta SeqKit locate (Shen *et al.* 2016), lo que permitió una falta de emparejamiento máxima de 2 pb de la secuencia diana de HindIII AAGCTT. El número de roturas que se solapan con estos sitios previstos se calculó usando bedtools intersect. Para

comparar con el experimento DSBcapture EcoRV (Lensing *et al.* 2016), se usó el mismo umbral de cobertura de ≥ 5 roturas por sitio para definir cada sitio de rotura inducida por HindIII.

5 **Detección y análisis de las DSB inducidas por AsiSI en células DivA**

Las posiciones de los sitios diana de AsiSI se calcularon de la misma manera que para HindIII, sin embargo, sin permitir faltas de emparejamiento y usando la secuencia GCGATCGC. Como las células DivA son femeninas, se eliminaron los sitios presentes en el cromosoma Y, dejando 1211 sitios para chr1-X. Para calcular rigurosamente las roturas genuinas inducidas por AsiSI, el sitio AsiSI de 8 pb se redujo a intervalos genómicos de 1 pb en las posiciones de rotura previstas. Esto redujo cada intervalo genómico de 8 pb a dos intervalos de 1 pb; en la posición 6 en la hebra positiva y en la posición 3 en la hebra negativa. Después, se calcularon solapamientos directos entre las posiciones de rotura de 1 pb y los sitios de rotura por AsiSI pronosticados, usando bedtools intersect. Se requirió una orientación de hebras coincidentes para que cada solapamiento se considerara un sitio de rotura genuino inducido por AsiSI.

15 **Esquema de análisis de eventos fuera de la diana de CRISPR**

Primero se predijeron dos conjuntos de posibles sitios fuera de la diana para EMX1 en hg19 usando la versión de línea de comando de Cas-OFFinder (Bae *et al.* 2014), lo que permitió hasta 6 faltas de emparejamiento en el espaciador y PAM canónico combinados para el primer conjunto, y hasta 7 faltas de emparejamiento para el segundo. A continuación, ambos conjuntos de secuencias predichas se filtraron en función del número de faltas de emparejamiento en la región de la semilla, definida como los 12 nucleótidos proximales al PAM. Cada conjunto se filtró hasta 2, 3, 4 y 5 faltas de emparejamiento en la semilla, generando un conjunto de 8 archivos con diferentes parámetros de filtrado de faltas de emparejamiento. Para definir las DSB inducidas por CRISPR, cada sitio previsto de 23 pb se redujo primero a un intervalo de 2 pb que flanquea la posición de rotura CRISPR esperada, 3 pb en dirección 5' del PAM. Entonces se calcularon los solapamientos entre estas regiones de rotura esperadas de 2 pb y las posiciones de rotura de 1 pb de INDUCE-seq usando bedtools intersect (Quinlan y Hall 2010), arrojando un conjunto de DSB identificadas en los sitios de rotura CRISPR esperados. Finalmente, las DSB que se superponen con los sitios CRISPR se filtraron según el número de faltas de emparejamiento del sitio y el número de roturas detectadas en el sitio. Los sitios que poseían faltas de emparejamiento $>n$ debían tener más de 1 solapamiento de DSB para ser retenidos como un sitio genuino fuera de la diana. Cada conjunto de solapamientos de roturas se filtró usando un valor de faltas de emparejamiento de >2 , >3 , >4 y >5 , lo que resultó en un total de 32 condiciones de filtro y conjuntos de datos fuera de la diana para cada muestra de INDUCE-seq.

35 **Cálculo de solapamientos entre los métodos de detección de sitios fuera de la diana de CRISPR**

Los sitios fuera de la diana de EMX1 se compararon con métodos alternativos CIRCLE-seq, Digenome-seq, GUIDE-seq, BLISS, y HTGTS. Se generaron archivos de intervalo del genoma para cada método de detección de sitios fuera de la diana respectivo. Los solapamientos de los sitios fuera de la diana de EMX1 detectados por cada método se calcularon usando bedtools intersect (Quinlan y Hall 2010).

40 **Validación de Amplicon-seq del resultado mutacional**

Las bibliotecas de ADN de secuenciación de amplicones se prepararon usando un panel personalizado de cebadores dependientes de RNasa-H (IDT) rhAmpSeq que flanquean los sitios fuera de la diana identificados por INDUCE-seq para EMX1. La PCR múltiple se llevó a cabo según las instrucciones del fabricante usando rhAmpSeq HotStart Master Mix 1, la mezcla de cebadores personalizada, y 10 ng de ADN genómico. Los productos de la PCR se purificaron usando perlas SPRI, y las secuencias de índice P5 y P7 de secuenciación de Illumina se incorporaron a través de una segunda PCR múltiple usando rhAmpSeq HotStart Master Mix 2. Las bibliotecas de secuenciación resultantes se reunieron y se secuenciaron usando una celda de flujo Illumina NextSeq 550 Mid Output con química de 2x 150 pb. Los resultados de la edición en los sitios dentro y fuera de la diana se determinaron usando el software CRISPResso (Pinello *et al.* 2016) v2.0.32, con los siguientes parámetros: CRISPRessoPooled -q30 -ignore_substitutions --max_paired_end_reads_overlap 151. Las frecuencias de indel se compararon usando CRISPRessoCompare.

55 **Resultados**

La medida de la rotura mediante INDUCE-seq se logra mediante una preparación de biblioteca sin PCR en dos etapas (Fig. 1 y 5). La primera etapa consiste en marcar las DSB preparadas en los extremos *in situ* mediante la ligación de un adaptador P5 modificado químicamente de longitud completa a los extremos de las DSB. En la etapa dos, el ADNg extraído, fragmentado y preparado en los extremos se liga usando un segundo adaptador P7 semifuncional modificado químicamente. Los fragmentos de ADN marcados con DSB resultantes que comprenden los adaptadores P5 y P7 semifuncionales pueden interactuar con la celda de flujo de Illumina, y posteriormente pueden secuenciarse mediante secuenciación de lectura única. Los fragmentos de ADN que no poseen el adaptador P5 permanecen no funcionales debido a que carecen de la secuencia necesaria para hibridarse con la celda de flujo. Esta metodología permite el enriquecimiento de secuencias de DSB funcionalmente marcadas y la eliminación de todos los demás fragmentos de ADN genómico, que de otro modo contribuirían al ruido del sistema. Evitar la amplificación de las roturas produce un resultado de secuenciación en el que una única lectura de secuenciación es equivalente a un único extremo de DSB

marcado (compárense las Fig. 5a y 5b, véase la Fig. 8a). Esta innovación permite la medida directa, y por tanto la cuantificación, de las DSB mediante secuenciación, lo que representa un avance importante en la medida precisa de las DSB en el genoma.

5 Después del marcaje de las roturas *in situ*, los métodos de detección de DSB disponibles actualmente, BLISS, DSBCapture y END-seq, emplean un protocolo de enriquecimiento para separar los fragmentos de ADN marcados con DSB del ADN genómico restante. A esto le sigue una preparación de biblioteca basada en PCR y una
 10 secuenciación (Fig. 5b). Este enfoque basado en amplificación da como resultado un resultado de secuenciación en el que una sola lectura no es equivalente a una sola rotura, en el que se requiere un protocolo de corrección de errores de PCR, tal como identificadores moleculares únicos (UMI), para intentar la cuantificación de DSB (Fig. 5b y Fig. 6b) (Yan *et al.* 2017). Es importante destacar que la nueva combinación de adaptador y preparación de biblioteca INDUCE-seq es compatible con cualquier protocolo de marcaje de DSB *in situ* publicado.

15 Para demostrar las características de la metodología INDUCE-seq, primero examinamos cómo INDUCE-seq mide las DSB en todo el genoma después de la inducción de las DSB definidas en células fijadas y permeabilizadas usando una endonucleasa de restricción HindIII de alta fidelidad. Este enfoque se ha usado anteriormente para comparar los métodos BLISS, END-seq y DSBCapture. Como se muestra en la Fig. 3a, INDUCE-seq detecta simultáneamente
 20 cientos de millones de DSB inducidas por HindIII altamente recurrentes, además de cientos de miles de DSB endógenas de nivel inferior dentro de la misma muestra, sin necesidad de ningún tipo de corrección de errores. Esto permite por primera vez la cuantificación y caracterización precisas de las DSB endógenas en el genoma. La Fig. 7 muestra un ejemplo de detección de DSB endógenas usando INDUCE-seq en células HEK293. En conjunto, estas observaciones muestran el rango dinámico notablemente amplio y la sensibilidad que se pueden lograr con INDUCE-seq. La Fig. 3b muestra que INDUCE-seq detecta el patrón de escisión de HindIII esperado, en el que dos bloques simétricos semisolapantes de lecturas de secuenciación directa e inversa se asignan a los sitios de escisión de HindIII conocidos
 25 en la hebra directa e inversa. Esto confirma que INDUCE-seq se puede usar para medir con precisión las estructuras finales de las DSB con una resolución de un solo nucleótido. Medimos un aumento drástico en las roturas por célula después del tratamiento de las células con HindIII, desde menos de 10 roturas endógenas por célula no tratada hasta más de 3.000 roturas inducidas por célula tratada. Esto demuestra que INDUCE-seq es capaz de medir cuantitativamente las roturas por celda en tres órdenes de magnitud (Fig. 3c). En comparación con un experimento equivalente que usa el método DSBCapture para detectar roturas inducidas por EcoRV, encontramos que una mayor proporción de las lecturas INDUCE-seq totales y alineadas se asignaron a sitios de restricción. Significativamente, el 96,7 % de las lecturas alineadas se asignaron a sitios de restricción, lo que representa una mejora del 25 % en la fidelidad de la detección de roturas con respecto a DSBCapture (Fig. 3d). Es importante destacar que INDUCE-seq usó 800 veces menos células que el experimento DSBCapture, e identificó una proporción similar de sitios de restricción HindIII (92,7 %) a la identificada por DSBCapture para EcoRV (93,7 %) (Fig. 3e). Además de los sitios HindIII
 35 diana que comprenden la secuencia AAGCTT, identificamos una cantidad sustancial de las DSB en una variedad de secuencias HindIII fuera de la diana, que difieren en una o dos bases que no coinciden (Fig. 3f). El número total de las DSB inducidas por HindIII medidas por INDUCE-seq osciló de ~150.000.000 en los sitios diana hasta sólo cinco DSB en el sitio fuera de la diana con la clasificación más baja. Por lo tanto, INDUCE-seq detecta cuantitativamente roturas en ocho órdenes de magnitud, lo que demuestra un rango dinámico de detección de roturas enormemente mejorado con respecto a otros métodos. También se observó una mayor sensibilidad cuando se llevaron a cabo experimentos similares que miden las DSB inducidas en sitios AsiSI en células DiVA vivas. A pesar de secuenciar 40 veces menos lecturas que un experimento de DSBCapture correspondiente y 23 veces menos lecturas que BLISS, INDUCE-seq detectó la presencia de roturas en 230 sitios AsiSI. Esto representa un aumento con respecto a los 214 sitios detectados por BLISS y los 121 sitios detectados por DSBCapture. Esto demuestra que INDUCE-seq es significativamente más sensible, eficiente, y por lo tanto rentable, que cualquier método de detección de rotura actual (Fig. 3g).

50 Una vez establecidas las características de la detección de roturas mediante INDUCE-seq, a continuación las aplicamos a la detección de las DSB dentro y fuera de la diana inducidas por CRISPR/Cas9 en el genoma. Este análisis es de importancia central en el perfil de seguridad para el desarrollo de terapias basadas en CRISPR. Después de la nucleofección mediante RNP de células HEK293 con el sgRNA de EMX1 ampliamente caracterizado, se determinaron sitios fuera de la diana a las 0, 7, 12, 24 y 30 horas después de la nucleofección, según lo definido por un proceso de análisis de datos personalizado (Fig. 4a, Fig. 8, y figura 9). La Fig. 3a también muestra el número de roturas detectadas en el sitio diana y en cada uno de los 60 sitios fuera de la diana para EMX1, con faltas de emparejamiento que varían de 2 pb a 6 pb en comparación con el sitio diana.

60 Este experimento revela un perfil de la cinética de inducción de roturas mediante la guía EMX1, ofreciendo información sobre el mecanismo del proceso de edición de CRISPR/Cas9. Como se muestra en la Figura 4b, la mayor parte de la actividad dentro y fuera de la diana se observó inmediatamente después de la nucleofección y durante las primeras etapas del transcurso del tiempo. Estos resultados demuestran la rapidez de la edición por este sgRNA. En comparación con las tecnologías existentes, encontramos que INDUCE-seq supera significativamente los métodos alternativos basados en células GUIDE-seq, BLISS y HTGTS, además de capturar varios sitios que anteriormente sólo se identificaban usando metodologías de descubrimiento *in vitro* fuera de la diana CIRCLE-seq y Digenoma-seq. Es importante destacar que INDUCE-seq también detecta nuevos sitios de rotura fuera de la diana que no se habían detectado previamente mediante ningún otro método (Fig. 4c y Fig. 10).

Finalmente, usando ADN de las mismas muestras, medimos el resultado de la edición en estos sitios de rotura dentro y fuera de la diana identificados por INDUCE-seq, mediante secuenciación de amplicones. La secuenciación de amplicones sólo puede detectar indels por encima de una tasa de descubrimiento falso de fondo del 0,1 %. Por lo tanto, se detectó evidencia de edición en el sitio diana, y sólo en cuatro de los 60 sitios fuera de la diana que se identificaron a lo largo del tiempo (Fig. 4d). Esta observación concuerda con un estudio previo idéntico, que identificó cinco sitios fuera de la diana con frecuencias de indel > 0,1 % 48 horas después de la nucleofección de células HEK293 con RNP de EMX1 (Fig. 4d, columna del extremo derecho, 48 h). Estos datos demuestran que INDUCE-seq puede detectar y cuantificar las DSB fuera de la diana inducidas por CRISPR con una sensibilidad mucho mayor que la posible para la detección de indels usando secuenciación de amplicones. Esta observación resalta la necesidad de métodos más sensibles para la detección de resultados de edición, con el fin de evaluar la seguridad de la edición del genoma. Curiosamente, un examen minucioso del patrón de rotura y el perfil de indel posterior tanto en el sitio diana como en los dos sitios fuera de la diana revela un patrón de escisión específico de sgRNA que se refleja y se correlaciona con el resultado de la edición (Fig. 11). Esto plantea la posibilidad de usar el patrón de DSB observado en los sitios de rotura inducidos por CRISPR para predecir el resultado final de la edición tanto en los sitios dentro como fuera de la diana.

Discusión

Hemos desarrollado una nueva metodología sin PCR para preparar bibliotecas de ADN para la secuenciación de próxima generación de las DSB en el genoma. Esto permite por primera vez medir directamente los extremos rotos en las células. Nuestro enfoque supera el problema de las malas relaciones señal-ruido para las DSB asociadas con la amplificación por PCR que normalmente se emplea en las preparaciones de bibliotecas NGS estándar. El novedoso diseño del adaptador INDUCE-seq esencialmente permite usar la celda de flujo de secuenciación para enriquecer las secuencias de DSB marcadas, evitando la necesidad de su amplificación. En cambio, la mejora en la relación señal-ruido se logra filtrando los extremos ruidosos generados durante la preparación de la muestra que no están asociados con las DSB fisiológicas genuinas que se encuentran en las células. Demostramos las características de INDUCE-seq para medir las DSB genómicas en una variedad de aplicaciones diferentes. Revelamos su capacidad para detectar de forma sensible y cuantitativa roturas inducidas por enzimas de restricción endógenas de bajo nivel y de alto nivel simultáneamente. Esto no había sido posible anteriormente sin la necesidad de métodos de corrección de errores complejos y costosos que tienen sus propias limitaciones e inconvenientes. Comparamos nuestros resultados con los métodos de detección de roturas disponibles actualmente para demostrar cómo los mejora, no sólo en términos de precisión y sensibilidad, sino también en términos de simplicidad, escalabilidad, facilidad de uso y rentabilidad. Todas estas son características esenciales de un ensayo que se puede usar para evaluar el perfil de seguridad de guías sintéticas para la edición del genoma CRISPR. Demostramos cómo se comporta INDUCE-seq en comparación con varios de los métodos de detección de DSB actuales para la detección de edición dentro y fuera de la diana mediante el sgRNA de EMX1. Revelamos que además de detectar muchos de los sitios fuera de la diana medidos acumulativamente por otros cinco métodos, INDUCE-seq también identifica un número significativo de sitios novedosos fuera de la diana que los métodos actuales no pueden detectar. Sugerimos que INDUCE-seq podría ser un método muy importante para la elaboración de perfiles de seguridad y el diseño de ARN guía sintético para el desarrollo futuro de la edición del genoma como modalidad terapéutica. Tales características incluyen mutaciones en todo el genoma, roturas y espacios en una sola hebra, así como otros tipos de daño en el ADN que pueden convertirse en roturas. El desarrollo de INDUCE-seq, y sus ensayos derivados, podría tener implicaciones importantes en una variedad de aplicaciones biomédicas diferentes.

Tabla 1.

| Par de cebadores | Secuencia (5'-3') | Identificador |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
| Primer par de cebadores (adaptador P5) | P-GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT AGATC TCGGTGGTCGCCGTATCATTC /3SpC3/ A*ATGATACGGCGACCACCGAG ATCT ACACTCTTCCCTACACGACGCTCTTCCGATC*T | SEQ ID NO: 1 SEQ ID NO: 2 |
| Segundo par de cebadores (adaptador P7) | P-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC/ 3SpC3 / C*AAGCAGAAGACGGCATACGAGAT[INDICE]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T C*AAGCAGAAGACGGCATACGAGAT CGTGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T | SEQ ID NO: 3 SEQ ID NO: 4 SEQ ID NO: 5 |

ES 2 984 253 T3

| Par de cebadores | Secuencia (5'-3') | Identificador |
|------------------|--------------------------------------------------------------------------------------|------------------|
| | C*AAGCAGAAGACGGCATAACGAGAT ACATCG GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 6 |
| | C*AAGCAGAAGACGGCATAACGAGAT GCCTAAG TGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 7 |
| | C*AAGCAGAAGACGGCATAACGAGAT TGGTCA GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 8 |
| | C*AAGCAGAAGACGGCATAACGAGAT CACTGT GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 9 |
| | C*AAGCAGAAGACGGCATAACGAGAT ATTGGC GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 10 |
| | C*AAGCAGAAGACGGCATAACGAGAT GATCTG GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 11 |
| | C*AAGCAGAAGACGGCATAACGAGAT TCAAGT GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 12 |
| | C*AAGCAGAAGACGGCATAACGAGAT CTGATC GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 13 |
| | C*AAGCAGAAGACGGCATAACGAGAT AAGCTA GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 14 |
| | C*AAGCAGAAGACGGCATAACGAGAT GTAGCC GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 15 |
| | C*AAGCAGAAGACGGCATAACGAGAT TACAAG GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T | SEQ ID NO: 16 |
| | C*AAGCAGAAGACGGCATAACGAGAT TGTGACT GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 17 |
| | C*AAGCAGAAGACGGCATAACGAGAT ACGGAAC TGTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 18 |
| | C*AAGCAGAAGACGGCATAACGAGAT TCTGACAT GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 19 |
| | C*AAGCAGAAGACGGCATAACGAGAT CGGGACGGG TGACTGGAGTTCAGACGTGTGCT CTTCCGATC*T | SEQ ID NO: 20 |
| | C*AAGCAGAAGACGGCATAACGAGAT GTGCGGAC GTGACTGGAGTTCAGACGTGTGCT CTTCCGATC*T | SEQ ID NO: 21 |
| | C*AAGCAGAAGACGGCATAACGAGAT CGTTTCA CGTGTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 22 |
| | C*AAGCAGAAGACGGCATAACGAGAT AAGGCCAC GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 23 |
| | C*AAGCAGAAGACGGCATAACGAGAT TCCGAAAC GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 24 |
| | C*AAGCAGAAGACGGCATAACGAGAT TACGTACG GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 25 |
| | C*AAGCAGAAGACGGCATAACGAGAT ATCCACTC GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 26 |
| | C*AAGCAGAAGACGGCATAACGAGAT ATCAGT GTGACTGGAGTTCAGACGTGTGCTCT TCCGATC*T | SEQ ID NO: 27 |
| | C*AAGCAGAAGACGGCATAACGAGAT AAAGGAAT GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T | SEQ ID NO: 28 |

| Par de cebadores | Secuencia (5'-3') | Identificador |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| | *Enlace de fosforotioato Resistencia a la actividad de exonucleasa | |
| | /3SpC3/ Espaciador C3 de fosforamidita (bloque covalente) Resistencia a la actividad de exonucleasa, ligasa y polimerasa 5'>3' | |
| | -P 5' Grupo fosfato Facilita la ligación 5' | |
| | ÍNDICE Secuencia de índice de secuenciación de Illumina Habilita la desmultiplexación de bibliotecas de secuenciación agrupadas | |
| | *T 3' trifosfato de desoxitimidina 'Cola T' con enlace de fosforotioato Proporciona sustrato para la ligación a fragmentos de ADN con 'cola A' | |

Tabla 2

| Modificación oligo (adaptador P5 o P7) | Actividad |
|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Enlaces de fosforotioato (extremos 5' y 3' del oligo) | Resiste la actividad de exonucleasa 5'>3' y 3'>5'. |
| Enlaces de fosforotioato (a lo largo de oligo) | Resiste la actividad de endonucleasa. |
| Fosforamidita espaciadora C3 (3') | Resiste la actividad de exonucleasa, ligasa, transferasa terminal, polimerasa 5'>3'. |
| Grupo 3' fosfato | Inhibe la degradación por algunas 3'-exonucleasas, puede usarse para bloquear el alargamiento por ADN polimerasas, y resiste la ligación. |
| 2'-O-Metil (2'OMe) | Los oligonucleótidos de ADN que incluyen esta modificación suelen ser 5 a 10 veces menos susceptibles a las ADNasas que el ADN no modificado. |
| dT invertida (3') | La dT invertida en el extremo 3' de un oligonucleótido conduce a un enlace 3'-3' que inhibirá la degradación por las 3' exonucleasas y el alargamiento por ADN polimerasas. |
| ddT 5' invertida | La ddT 5' invertida en el extremo 5' de un oligonucleótido resiste la ligación, y puede proporcionar resistencia a algunas formas de exonucleasa. |
| Bases de ácido nucleico bloqueadas | Aumenta la temperatura de fusión del oligo, y previene la hibridación espuria, aumenta la especificidad de unión. |
| Didesoxicitidina (ddC) (3') | Terminador de cadena 3' que evita el alargamiento de 3' por polimerasas. |

5

Referencias

- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Disponible en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> [Acceso: 2019].
- 10 Bae, S., Park, J. y Kim, JS (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30(10):1473-1475.
- 15 Krueger, F. (2015). Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Disponible en: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Acceso: 2019].
- 20 Lensing, S. V., Marsico, G., Hansel-Hertsch, R., Lam, E. Y., Tannahill, D. y Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. *Nature Methods* 13(10):855-+.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
- 25 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Genome Project Data, P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E. y Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology* 34(7):695-697.

5 Quinlan, A. R. y Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.

10 Shen, W., Le, S., Li, Y. y Hu, F. Q. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *Plos One* 11(10).

Yan, W. X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M. W., Kallas, T., ... Crosetto, N. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nature Communications* 8.

REIVINDICACIONES

1. Un método de preparación de muestras para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido nucleico, en el que la preparación comprende modificar ácidos nucleicos asociados a DSB para que sean:
- 5 (i) adecuados para unirse a un sustrato que comprende un primer cebador inmovilizado, y
(ii) adecuados para amplificación, en el que la amplificación comprende el uso de dicho primer cebador inmovilizado y un segundo cebador; comprendiendo el método:
- 10 a) proporcionar una muestra que comprende una pluralidad de ácidos nucleicos;
b) exponer la pluralidad de ácidos nucleicos a un primer adaptador en condiciones propicias para la ligación, en el que el primer adaptador comprende un oligonucleótido capaz de ligarse a un extremo 3' de una hebra de una DSB y que comprende una secuencia que es capaz de unirse al primer cebador inmovilizado mediante hibridación;
c) fragmentar la pluralidad de ácidos nucleicos; y
d) exponer la pluralidad de ácidos nucleicos a un segundo adaptador en condiciones propicias para la ligación, en el que el segundo adaptador comprende un oligonucleótido capaz de ligarse a un extremo 5' de una hebra en una rotura inducida por la fragmentación, pero no es capaz de ligarse al primer adaptador, y que no comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación, y que comprende una secuencia idéntica a una región del segundo cebador.
- 15 2. El método de la reivindicación 1, en el que el segundo cebador está inmovilizado al sustrato.
- 20 3. El método de la reivindicación 1 o la reivindicación 2, en el que la etapa b) es:
exponer la pluralidad de ácidos nucleicos a un primer par de adaptadores en condiciones propicias para la ligación, en el que el primer par de adaptadores es capaz de ligarse a al menos un extremo 3' de una hebra de una DSB, y en el que el primer par de adaptadores comprende un primer y un segundo oligonucleótidos que son al menos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 3' y comprende una secuencia que es capaz de unirse a un cebador inmovilizado al sustrato mediante hibridación; y
25 en el que la etapa d) es:
exponer la pluralidad de ácidos nucleicos a un segundo par de adaptadores en condiciones propicias para la ligación, en el que el segundo par de adaptadores es capaz de ligarse a al menos un extremo 5' de una hebra en una rotura inducida por fragmentación pero no es capaz de ligarse al primer oligonucleótido del primer par de adaptadores, en el que el segundo adaptador comprende un primer y un segundo oligonucleótidos parcialmente complementarios, y el primer oligonucleótido se puede ligar a un extremo 5' y comprende una secuencia idéntica a una región de un segundo cebador, y el segundo oligonucleótido no comprende una secuencia que es complementaria a dicha secuencia idéntica a una región del segundo cebador.
- 30 4. El método de una cualquiera de las reivindicaciones 1 a 3, en el que el segundo par de adaptadores comprende:
un primer oligonucleótido que comprende una secuencia según according to CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO: 32), y un segundo oligonucleótido que comprende una secuencia que no comprende una secuencia de más de 5, 10, 15, o 20 bases, o que no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); o
un primer oligonucleótido que comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34), y un segundo oligonucleótido que comprende una secuencia que no comprende una secuencia de más de 5, 10, o 15 bases, o que no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31).
- 35 5. El método de la reivindicación 3 o la reivindicación 4, en el que el primer y/o el segundo oligonucleótido del primer par de adaptadores comprende una característica de protección de 3' y/o 5'; y/o en el que el primer y/o el segundo oligonucleótido del segundo par de adaptadores comprende una característica de protección de 3' y/o 5'; opcionalmente en el que el segundo adaptador no puede ligarse al primer adaptador debido a la presencia de una modificación 3' del primer adaptador.
- 40 6. El método de una cualquiera de las reivindicaciones 1 a 5, en el que el oligonucleótido del segundo adaptador que se puede ligar a un extremo 5' comprende una secuencia idéntica a 5, 10, 15, 20, 21, 24, o más bases de un cebador inmovilizado.
- 45 7. El método de una cualquiera de las reivindicaciones 1 a 6, que comprende además poner en contacto la pluralidad de ácidos nucleicos con el sustrato que comprende cebadores inmovilizados, en condiciones adecuadas para la hibridación de los cebadores inmovilizados con ácidos nucleicos complementarios.
- 50 8. El método de la reivindicación 7, que comprende además obtener información de secuencia para cualesquiera ácidos nucleicos hibridados con el sustrato.
- 55 9. El método de una cualquiera de las reivindicaciones 1 a 8, en el que dicha muestra que comprende una pluralidad de ácidos nucleicos es ADNg.
- 60 65

10. El método de una cualquiera de las reivindicaciones 1 a 9, en el que:
 las etapas se llevan a cabo en el orden a), b), c), y después d); o
 las etapas se llevan a cabo en el orden a), c), d), y después b); en el que la muestra está expuesta a condiciones
 capaces de causar o que se sospecha que pueden causar una DSB entre las etapas d) y b); opcionalmente en el que
 la muestra se expone a condiciones capaces de provocar una DSB en una característica de interés en la muestra de
 ácido nucleico.
11. Un kit para la preparación de muestras para identificar las DSB en una muestra de ADN, en el que el kit comprende
 pares de oligonucleótidos que están configurados para ser adecuados para actuar como el primer y segundo adaptador
 cuando se llevan a cabo los métodos de una cualquiera de las reivindicaciones 1 a 10, comprendiendo el kit
 i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la
 ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según
 TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que es complementario a dicho primer
 oligonucleótido del primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una
 característica de protección de 3' y/o 5'; y
 ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, 15, o
 20 bases, o no comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un
 segundo oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido
 a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según
 CAAGCAGAAGACGGCATAACGAGAT (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos
 comprenden, respectivamente, una característica de protección de 3' y/o 5'; o
 i) un primer par de oligonucleótidos, el primero de los cuales comprende una característica de unión 5' que permite la
 ligación de dicho oligonucleótido a una hebra de un ácido nucleico bicatenario, y comprende una secuencia según
 ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo oligonucleótido que es complementario a dicho
 primer oligonucleótido del primer par; y en el que uno o ambos de dichos oligonucleótidos comprenden,
 respectivamente, una característica de protección de 3' y/o 5'; y
 ii) un segundo par de oligonucleótidos, el primero de los cuales no comprende una secuencia de más de 5, 10, o 15
 bases, o no comprende las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo
 oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una
 hebra de un ácido nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID
 NO: 34); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de
 protección de 3' y/o 5'.
12. El kit según la reivindicación 11, en el que:
 (a) el primer oligonucleótido del primer par y/o del segundo par de oligonucleótidos comprende un sitio de hibridación
 al que se puede unir un primer cebador de secuenciación; y/o
 (b) dichos primer y segundo oligonucleótidos de la parte i) y/o parte ii) también comprenden una característica de
 índice que es una secuencia particular de nucleótidos que permite determinar el origen de las muestras reunidas.
13. El kit según la reivindicación 12, en el que dicho kit comprende además:
 al menos un cebador que se une al primer y/o segundo sitio de hibridación con fines de secuenciación; y/o
 agentes de fragmentación.
14. Un adaptador de doble hebra para identificar roturas de doble hebra (DSB) de ADN en una muestra de ácido
 nucleico, en el que el adaptador de doble hebra está configurado para ser adecuado para actuar como el segundo
 adaptador cuando se llevan a cabo los métodos de una cualquiera de las reivindicaciones 1 a 10, comprendiendo el
 adaptador de doble hebra:
 (a) una primera hebra de oligonucleótidos no comprende una secuencia de más de 5, 10, 15, o 20 bases, o no
 comprende las 24 bases, de la secuencia ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 30); y un segundo
 oligonucleótido que comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una
 hebra de un ácido nucleico bicatenario, y comprende una secuencia según CAAGCAGAAGACGGCATAACGAGAT
 (SEQ ID NO: 32); y en el que uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una
 característica de protección de 3' y/o 5'; o
 (b) una primera hebra de oligonucleótidos no comprende una secuencia de más de 5, 10, o 15 bases, o no comprende
 las 20 bases, de la secuencia TCGGTGGTCGCCGTATCATT (SEQ ID NO: 31); y un segundo oligonucleótido que
 comprende una característica de unión 3' que permite la ligación de dicho oligonucleótido a una hebra de un ácido
 nucleico bicatenario, y comprende una secuencia según AATGATACGGCGACCACCGA (SEQ ID NO: 34); y en el que
 uno o ambos de dichos oligonucleótidos comprenden, respectivamente, una característica de protección de 3' y/o 5'.

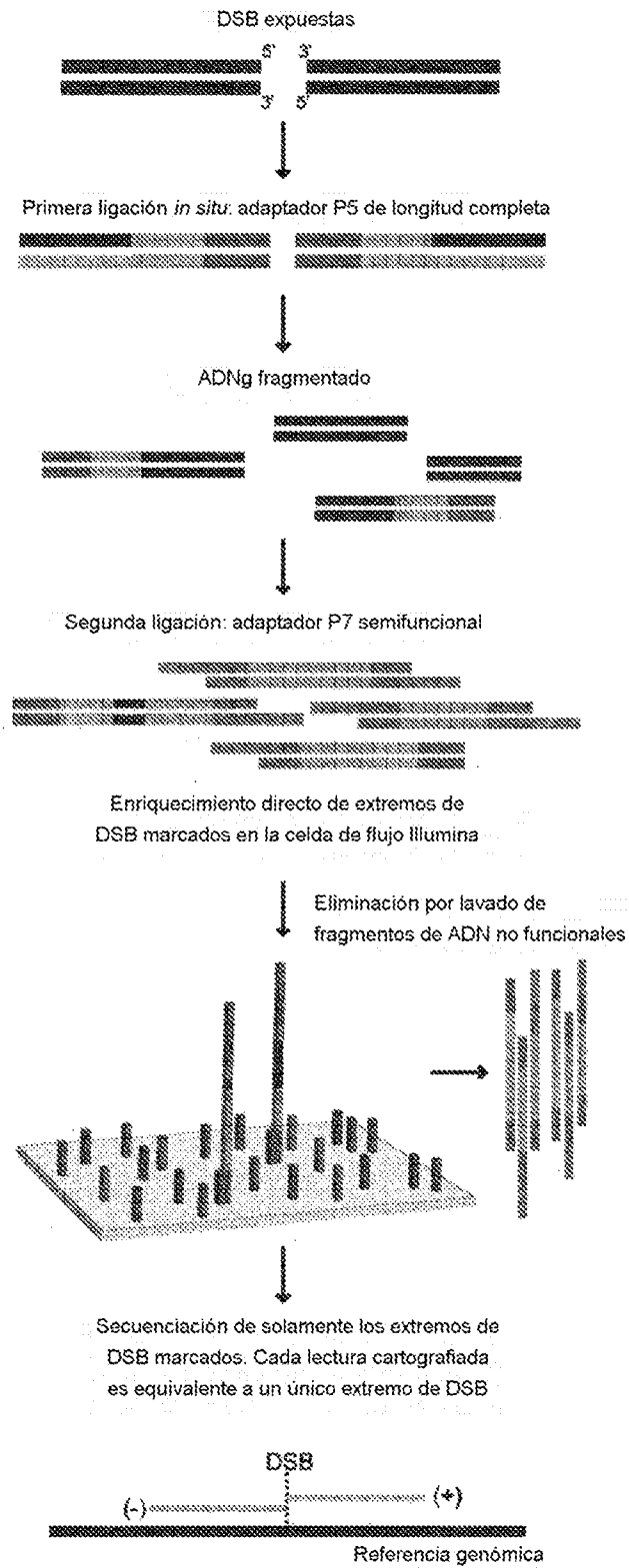


Figura 1

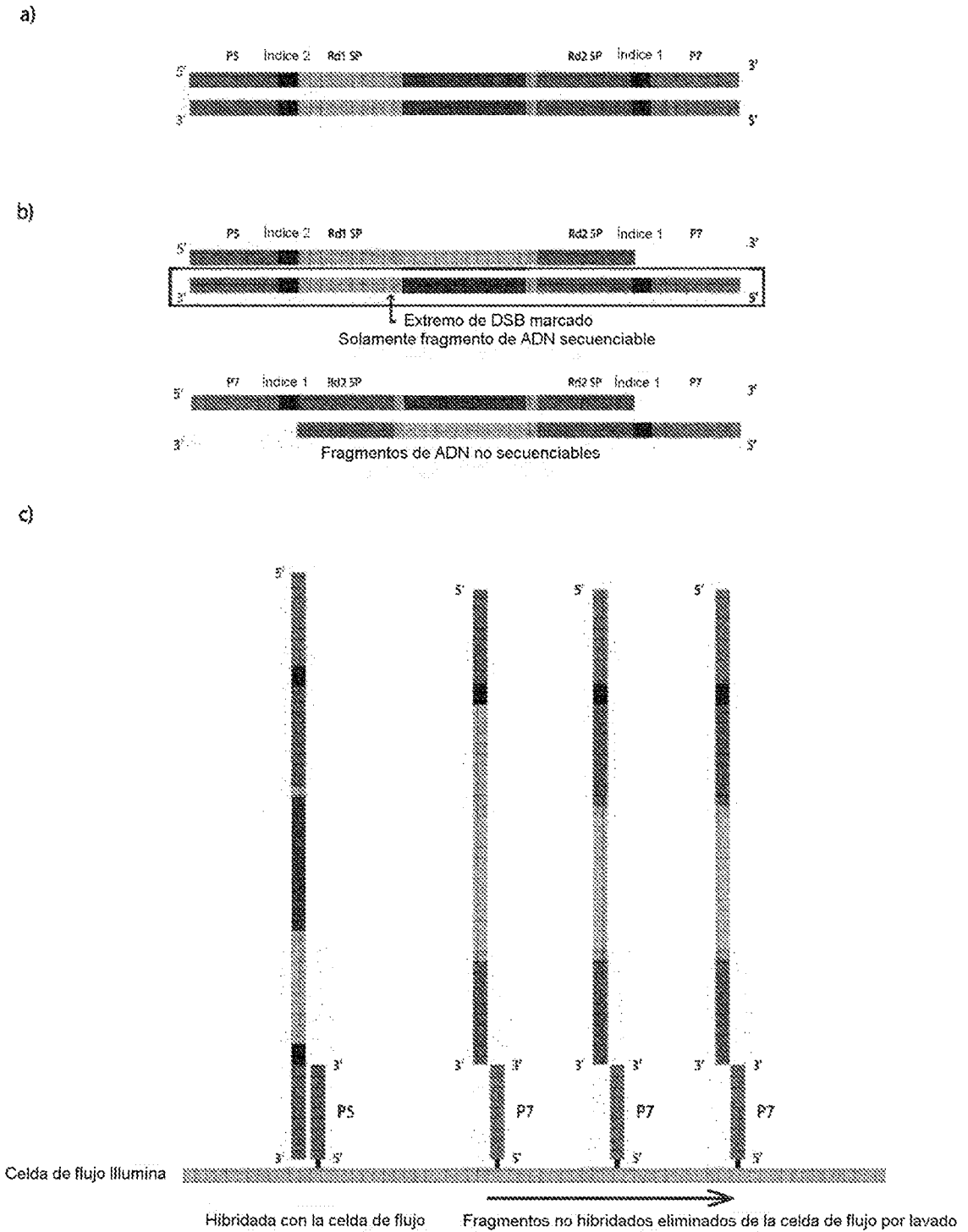


Figura 2

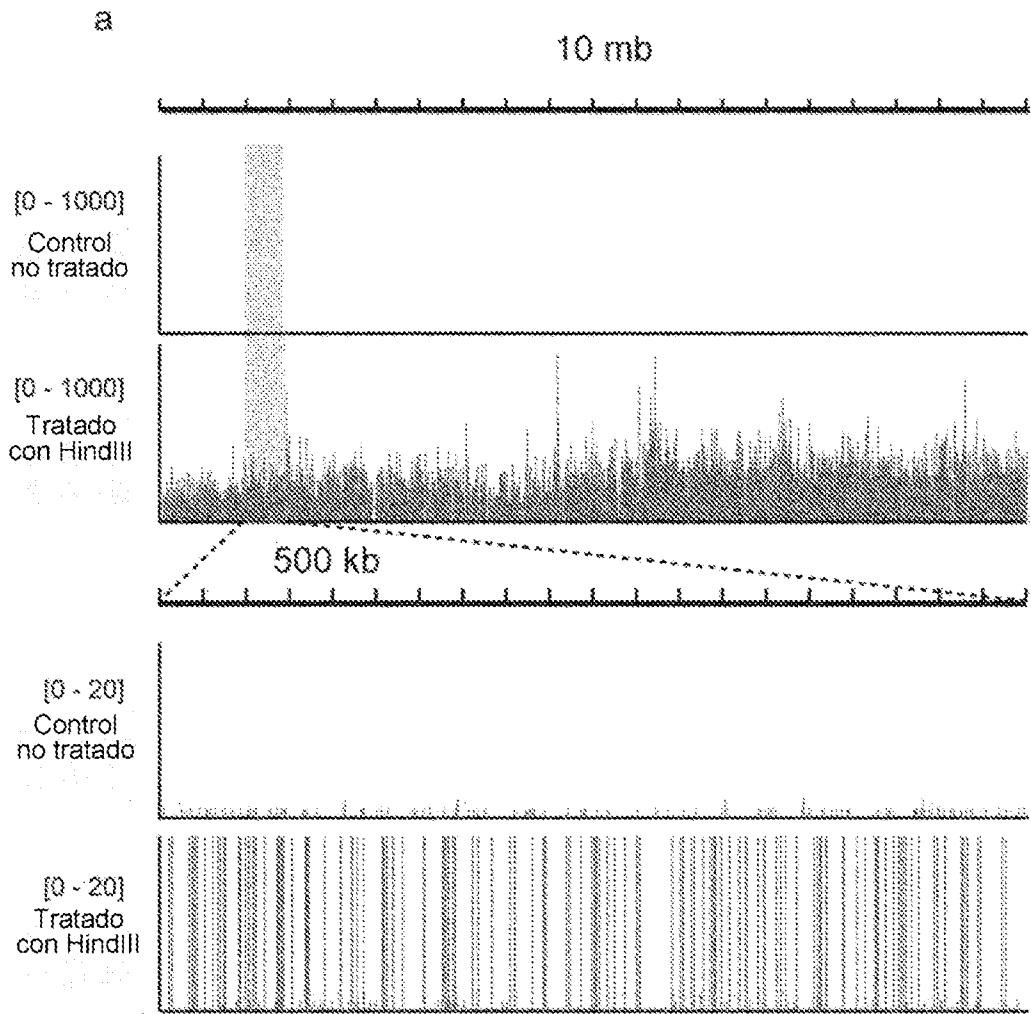


Figura 3

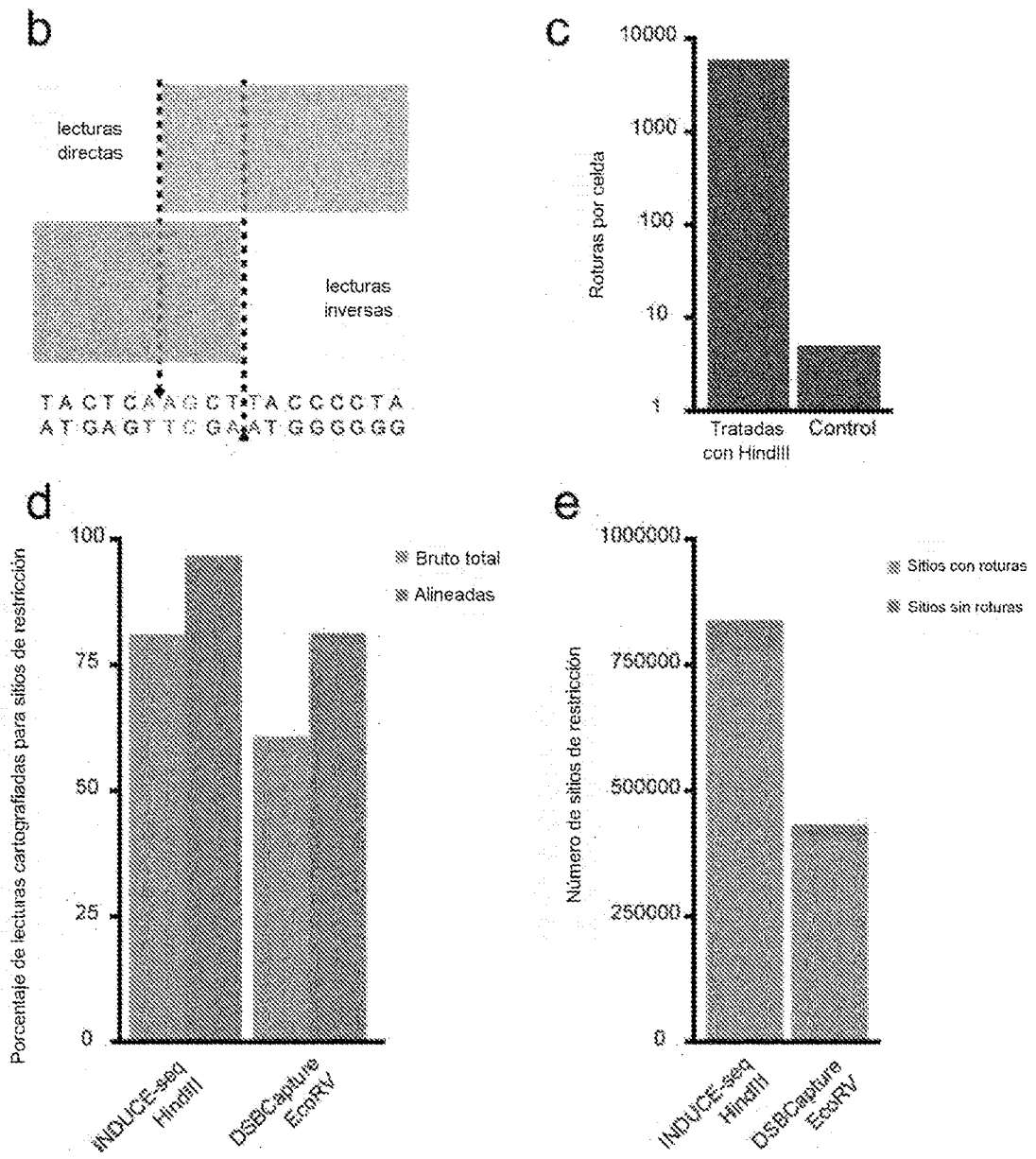


Figura 3 (cont.)

f Número de roturas detectadas en cada sitio de restricción

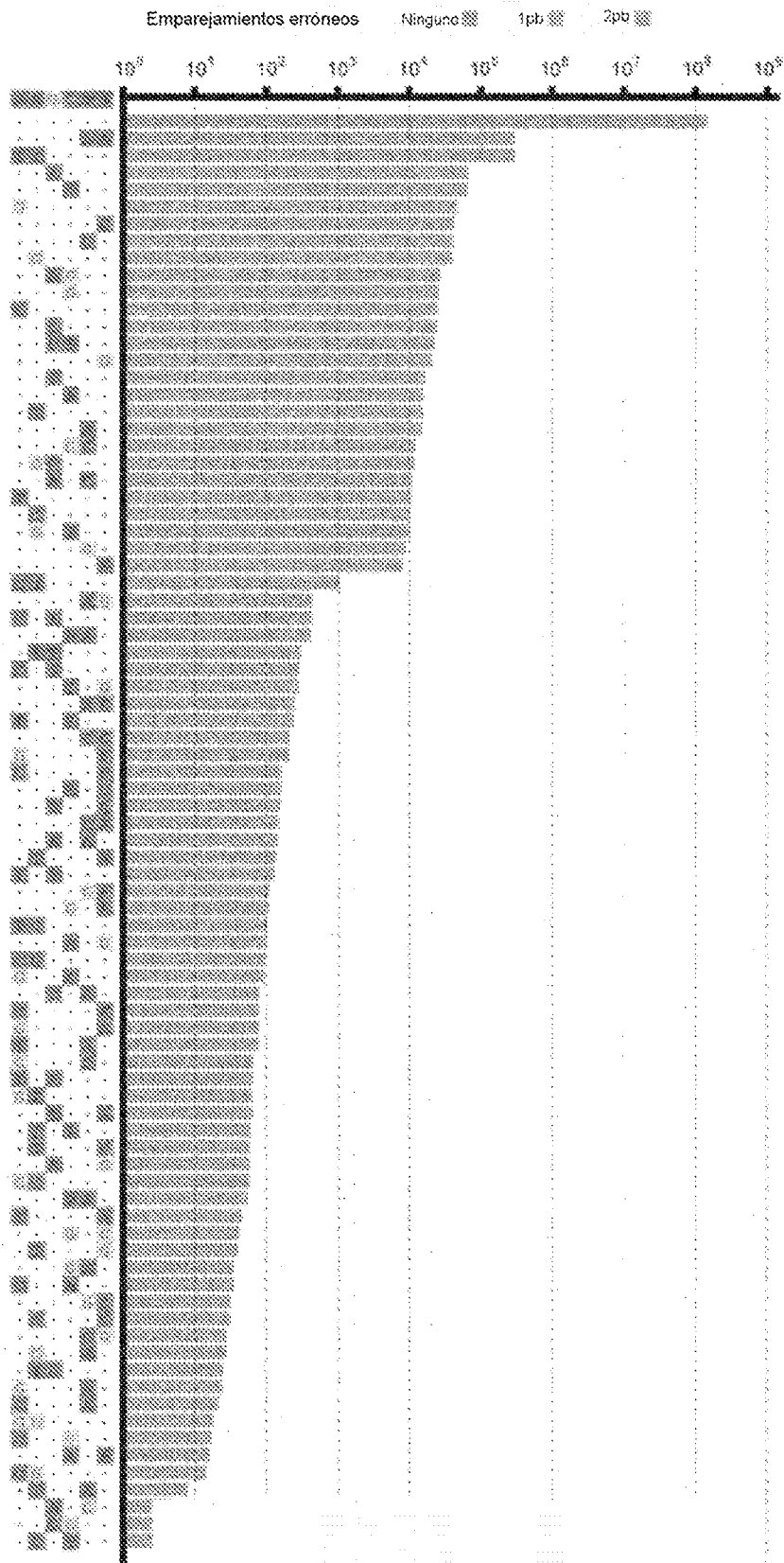


Figura 3 (cont.)

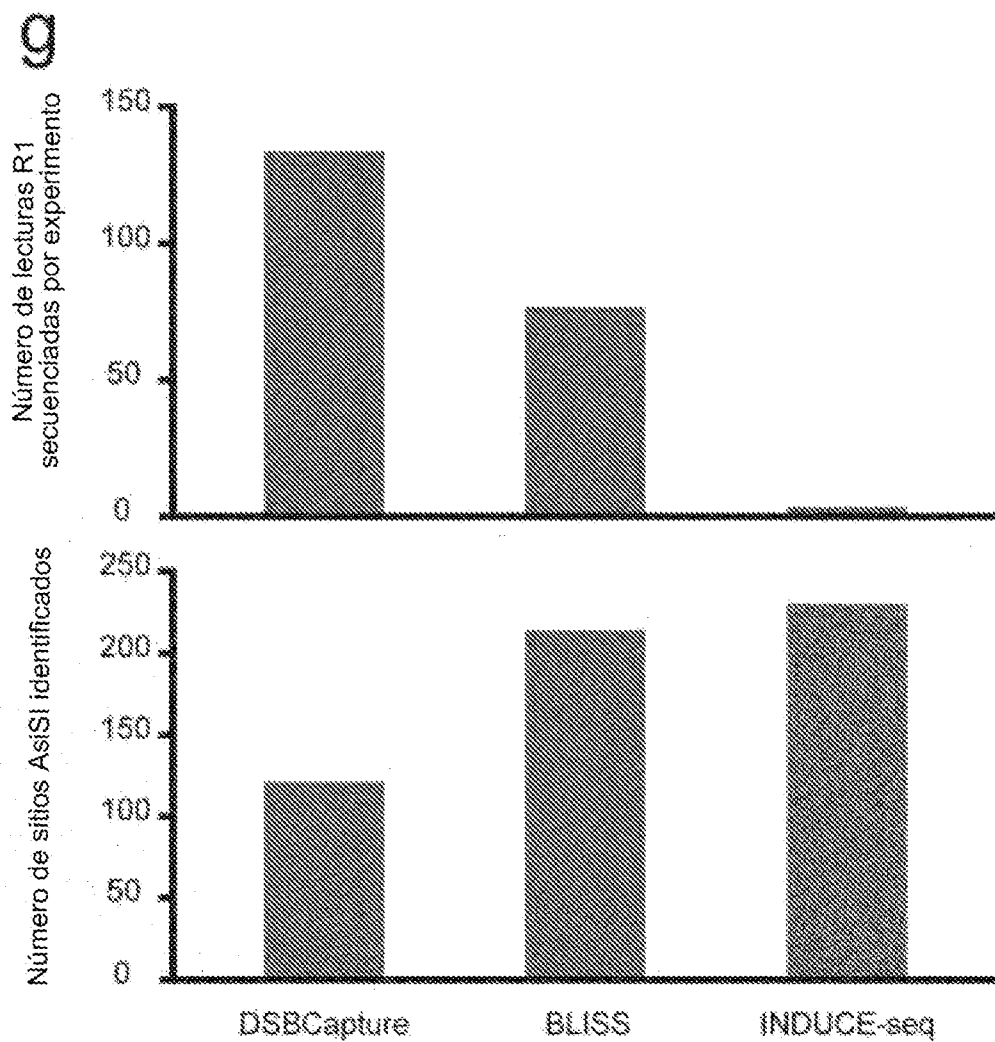
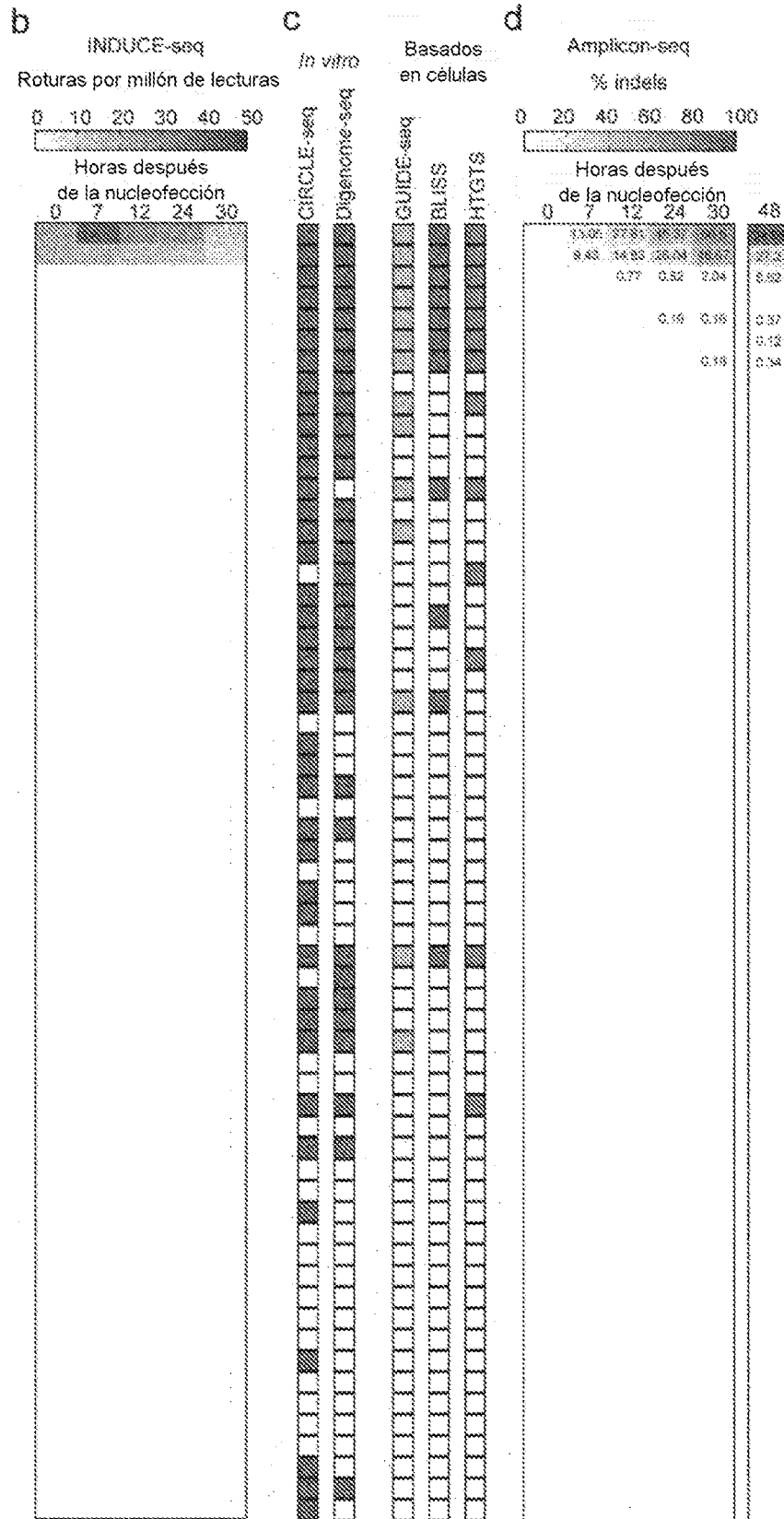
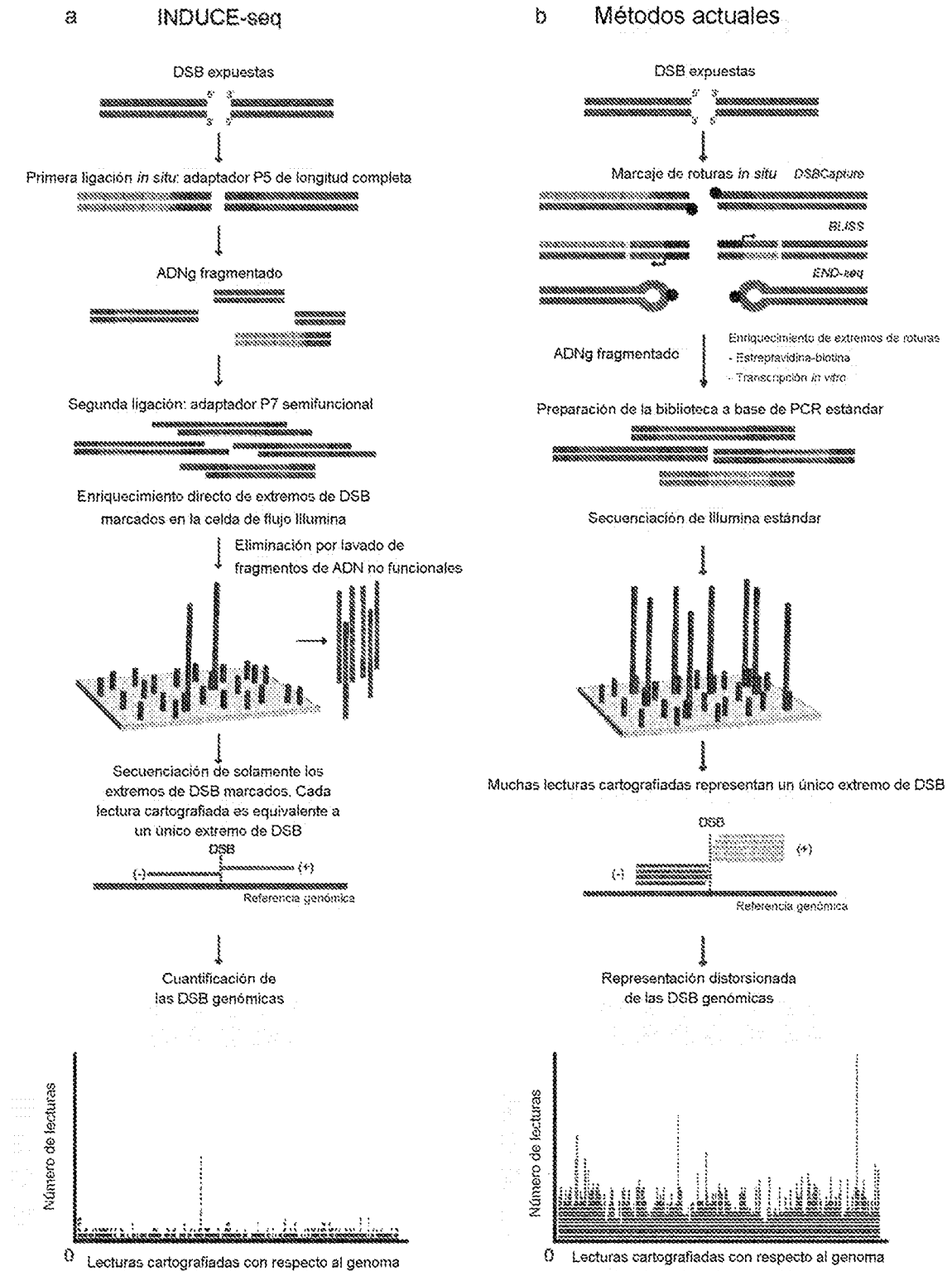


Figura 3 (cont.)

Figura 4
(cont.)





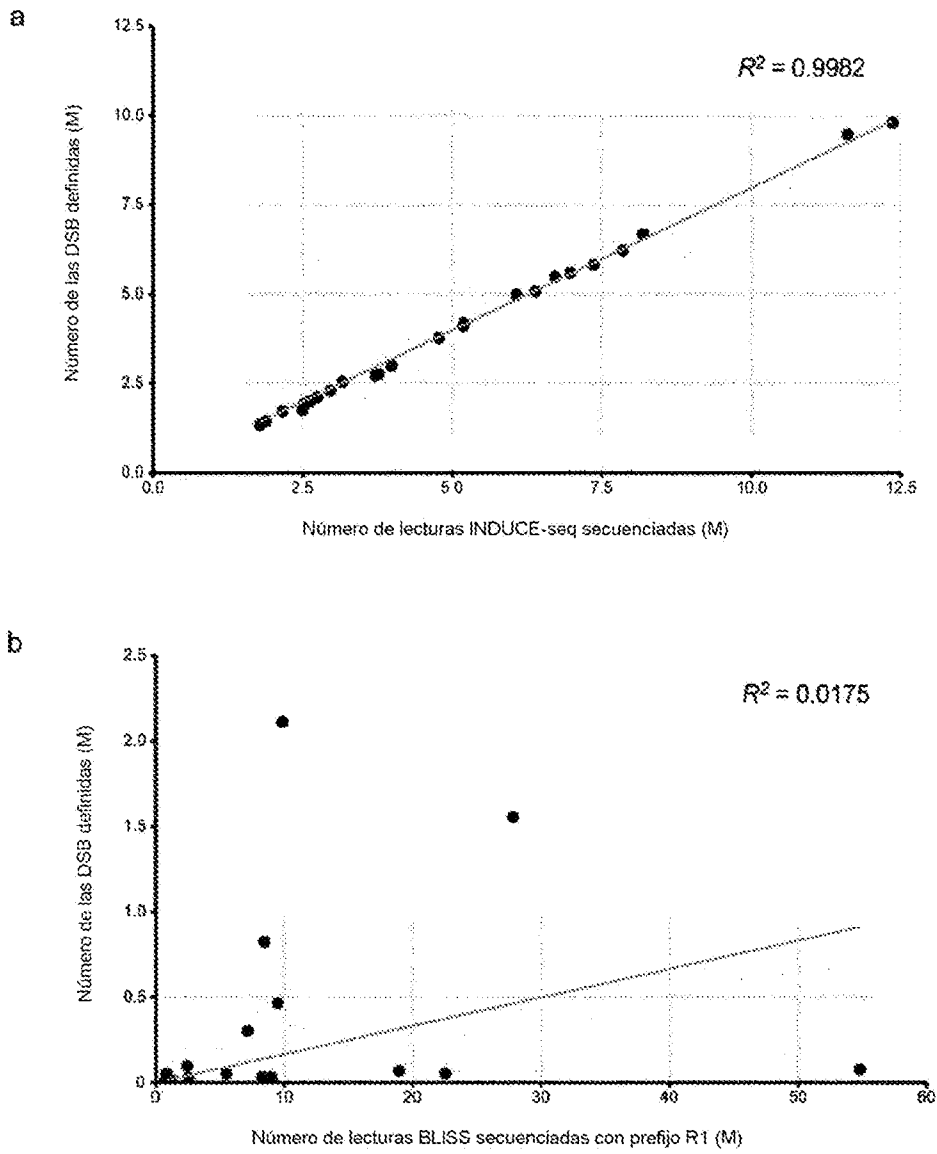


Figura 6

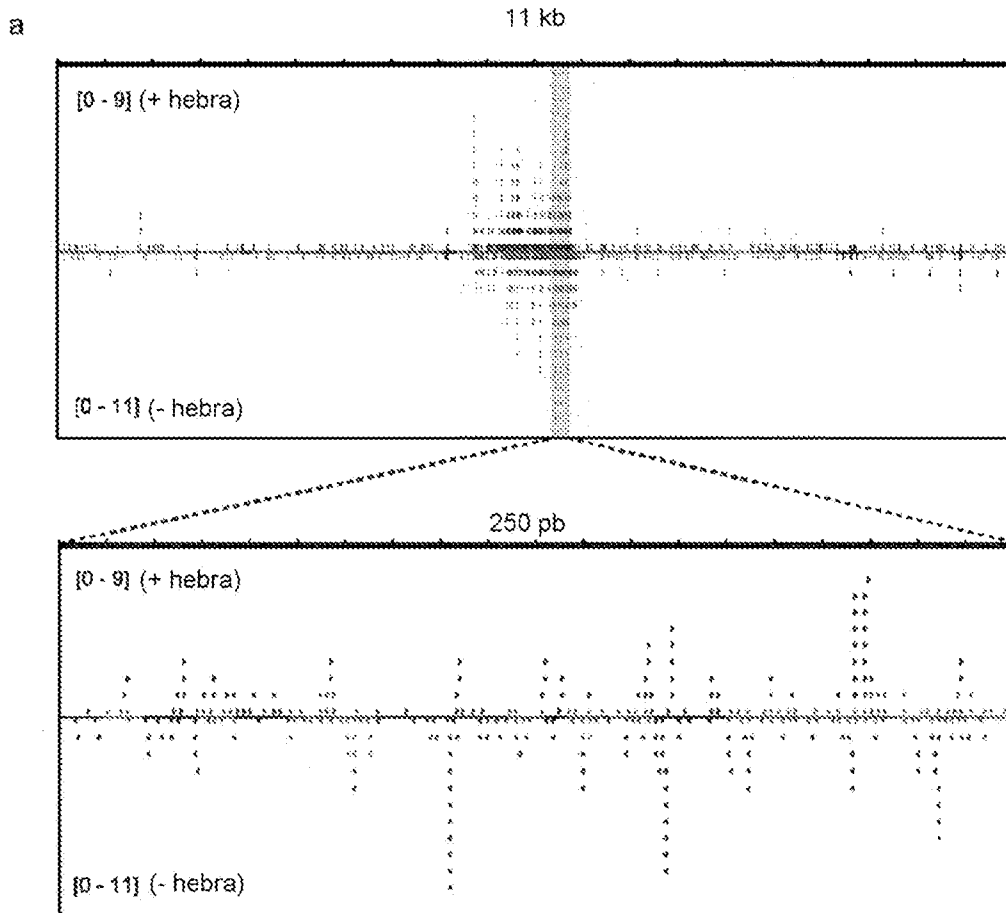


Figura 7

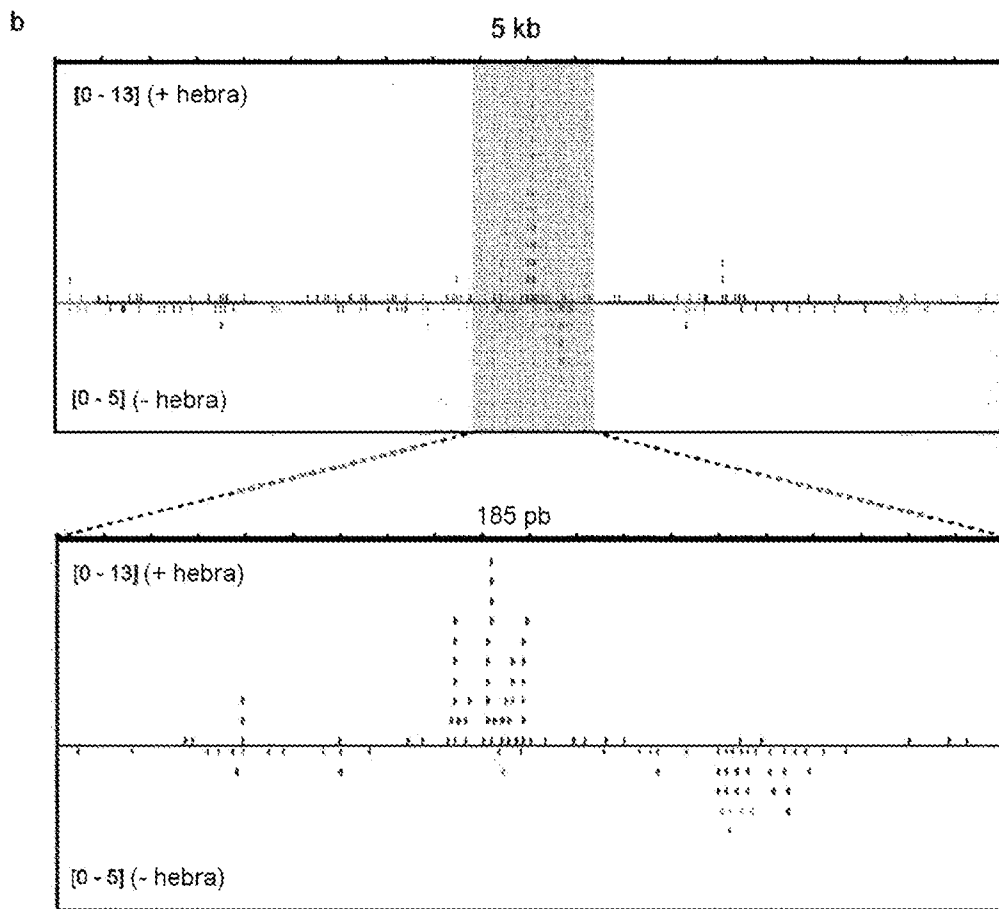


Figura 7 (cont.)

Esquema de descubrimiento de eventos fuera de diana de CRISPR

Prediga eventos fuera de diana de CRISPR para EMX1 en base a la secuencia (Cas-OFFinder) que permite hasta 7 pb en el Espaciador y PAM combinados.

Filtre eventos fuera de diana con más de 4 pb de emparejamientos erróneos en la región de siembra

Reduzca los sitios fuera de diana predichos a la posición de rotura esperada de 2 pb

Calcule el número de solapamientos entre posiciones de rotura de 1 pb INDUCE-seq y posiciones fuera de diana predichas de 2 pb

Retenga los eventos fuera de diana identificados en base a los emparejamientos erróneos y al número de roturas. Elimine los sitios con ≥ 5 pb y < 2 solapamientos.

Los sitios que quedan se definen como las DSB inducidas por CRISPR genuinas

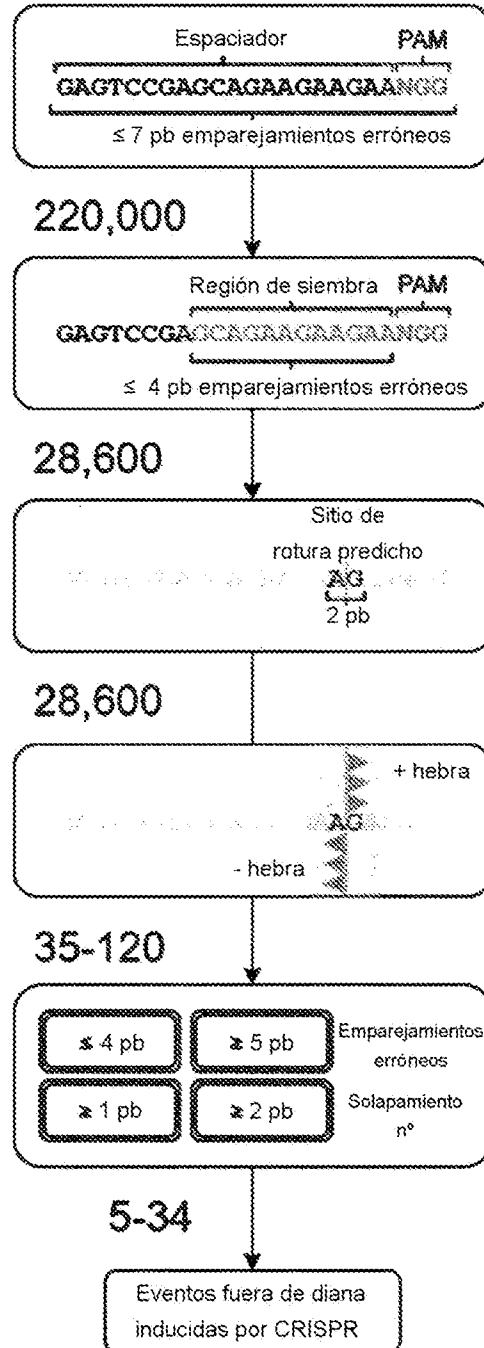


Figura 8

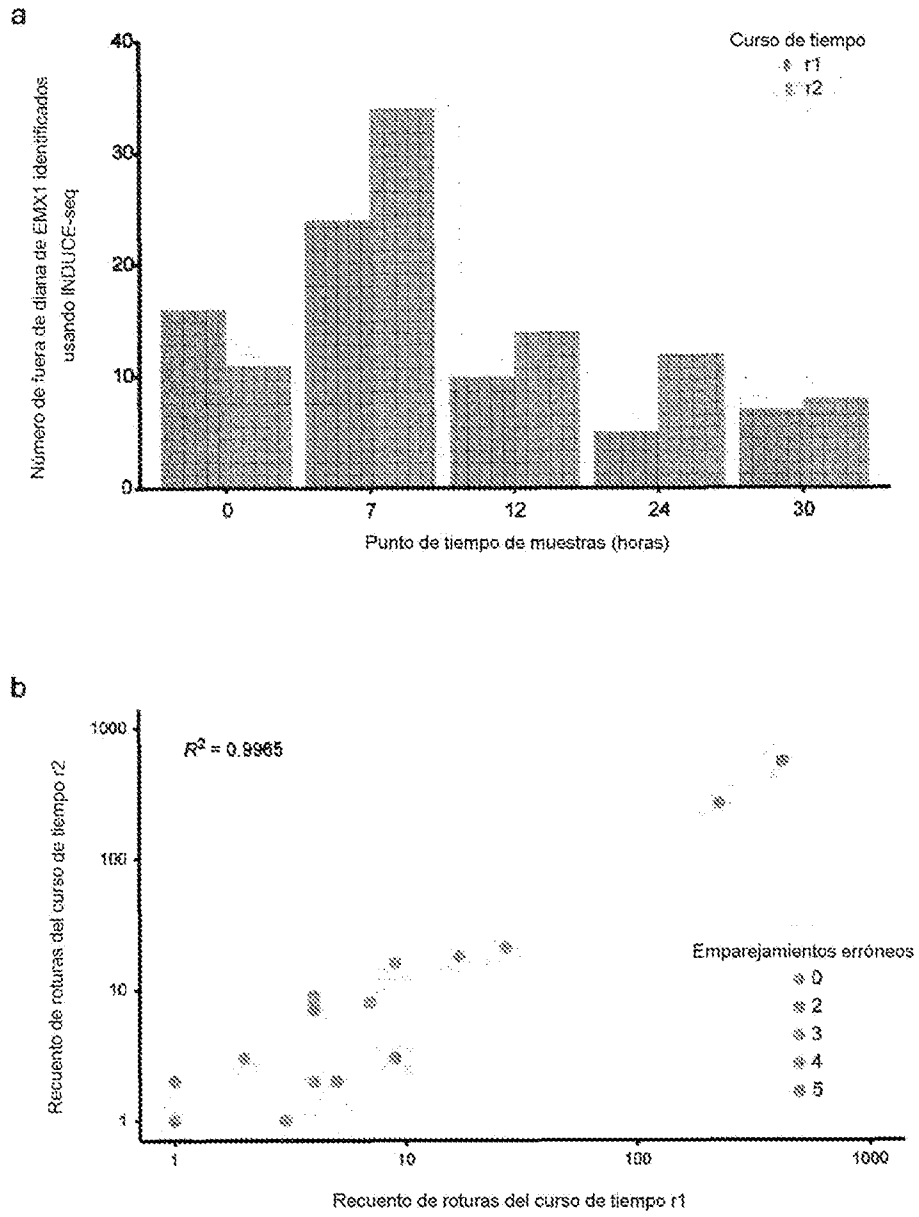


Figura 9

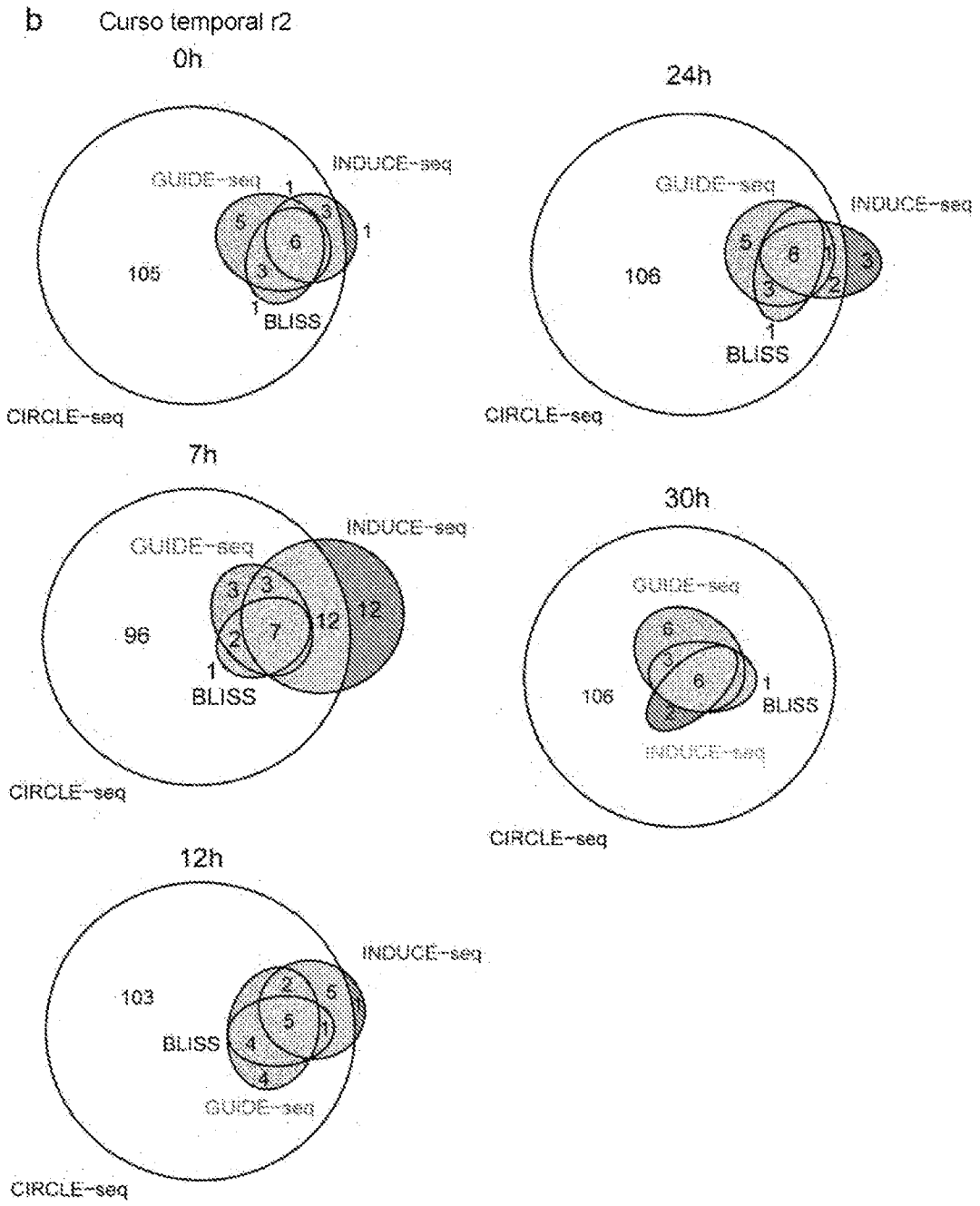
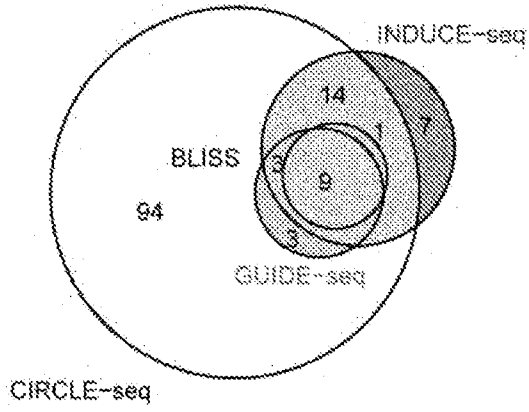
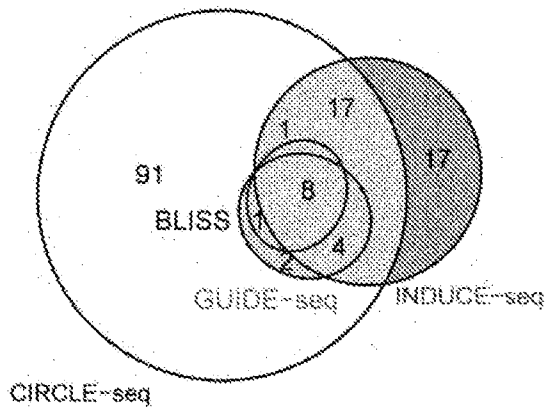


Figura 10 (cont.)

c Curso temporal r1 0h-30h



d Curso temporal r2 0h-30h



e Curso temporal r1 y r2 0h-30h

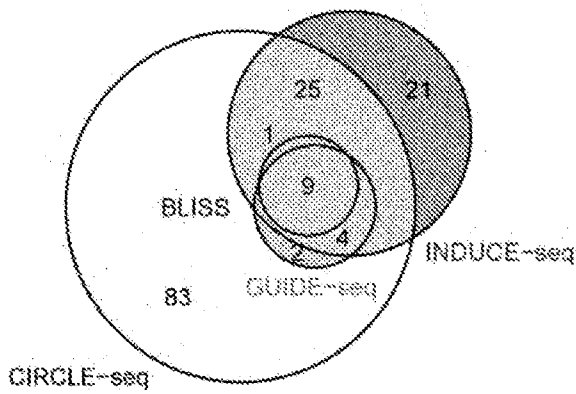


Figura 10 (cont.)

b

EMX1 OT-1

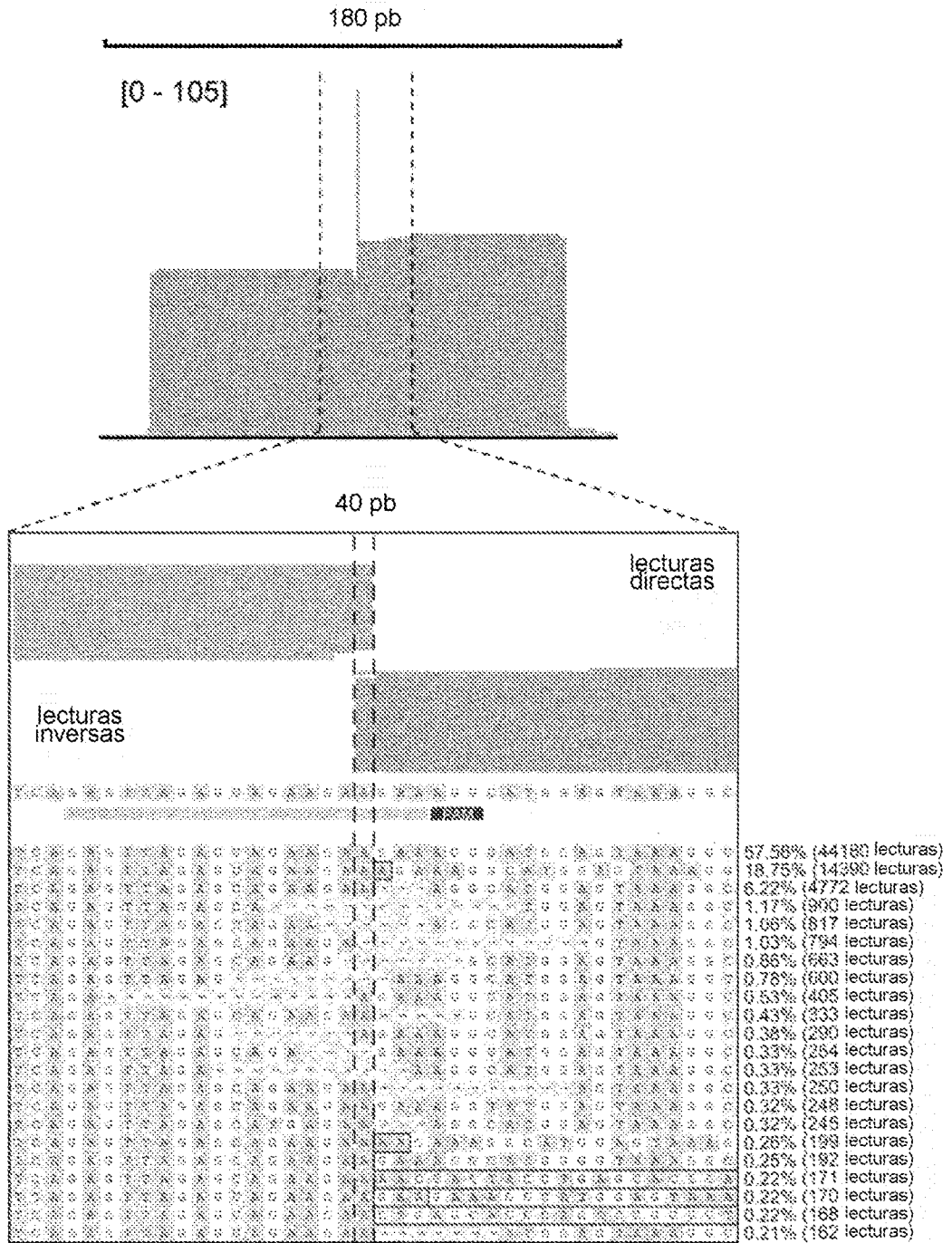


Figura 11 (cont.)

c

EMX1 OT-2

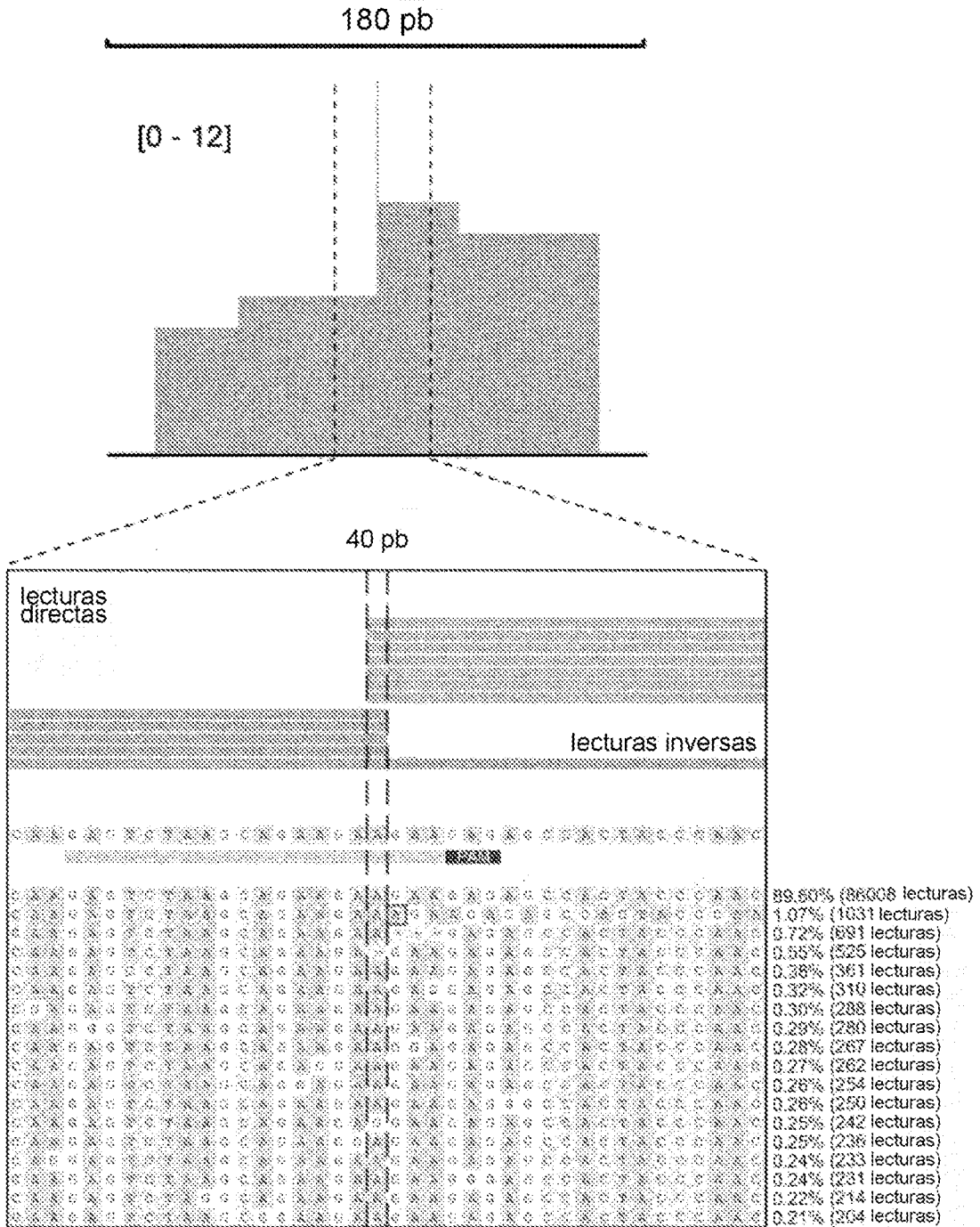


Figura 11 (cont.)

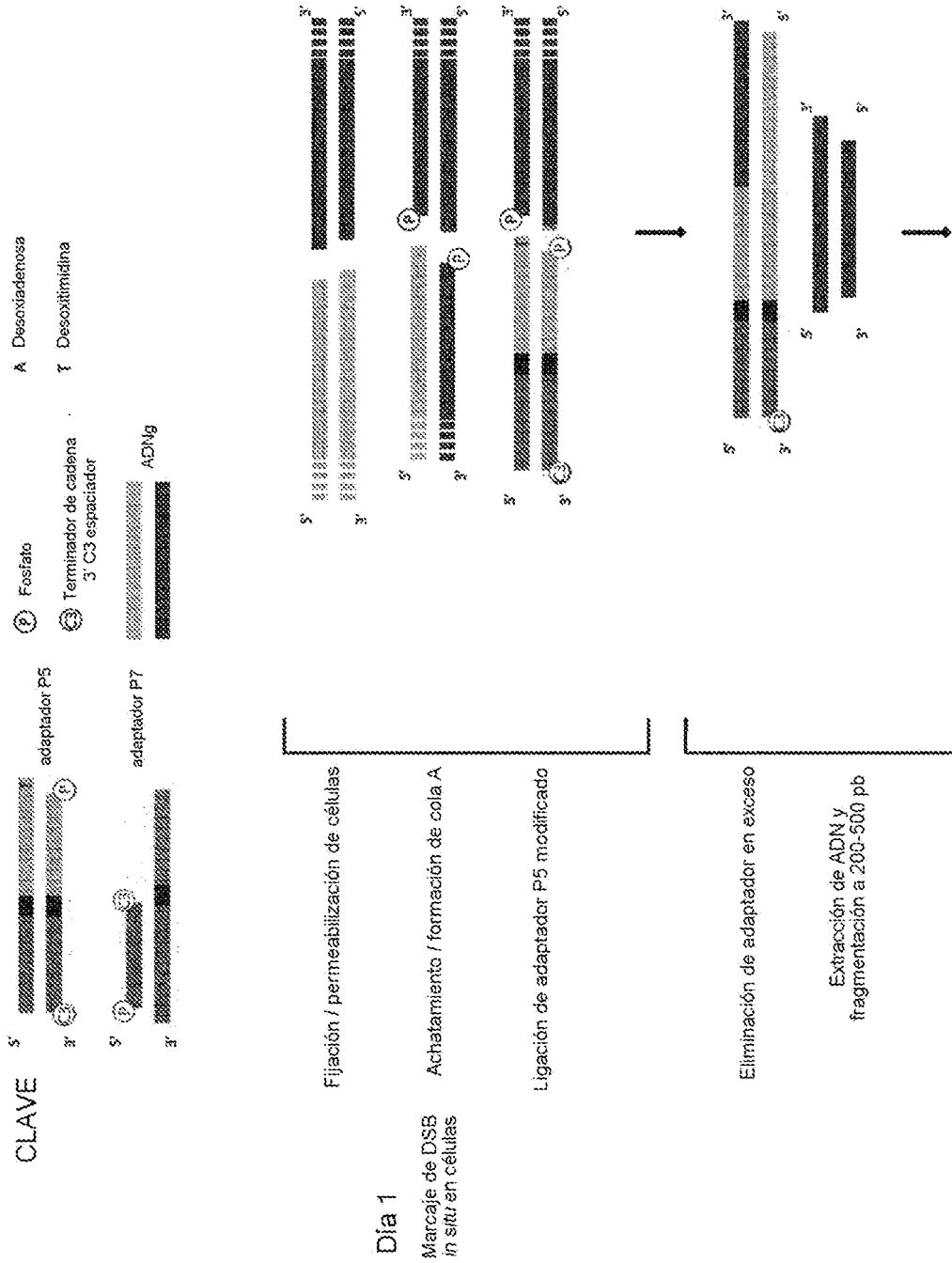


Figura 12

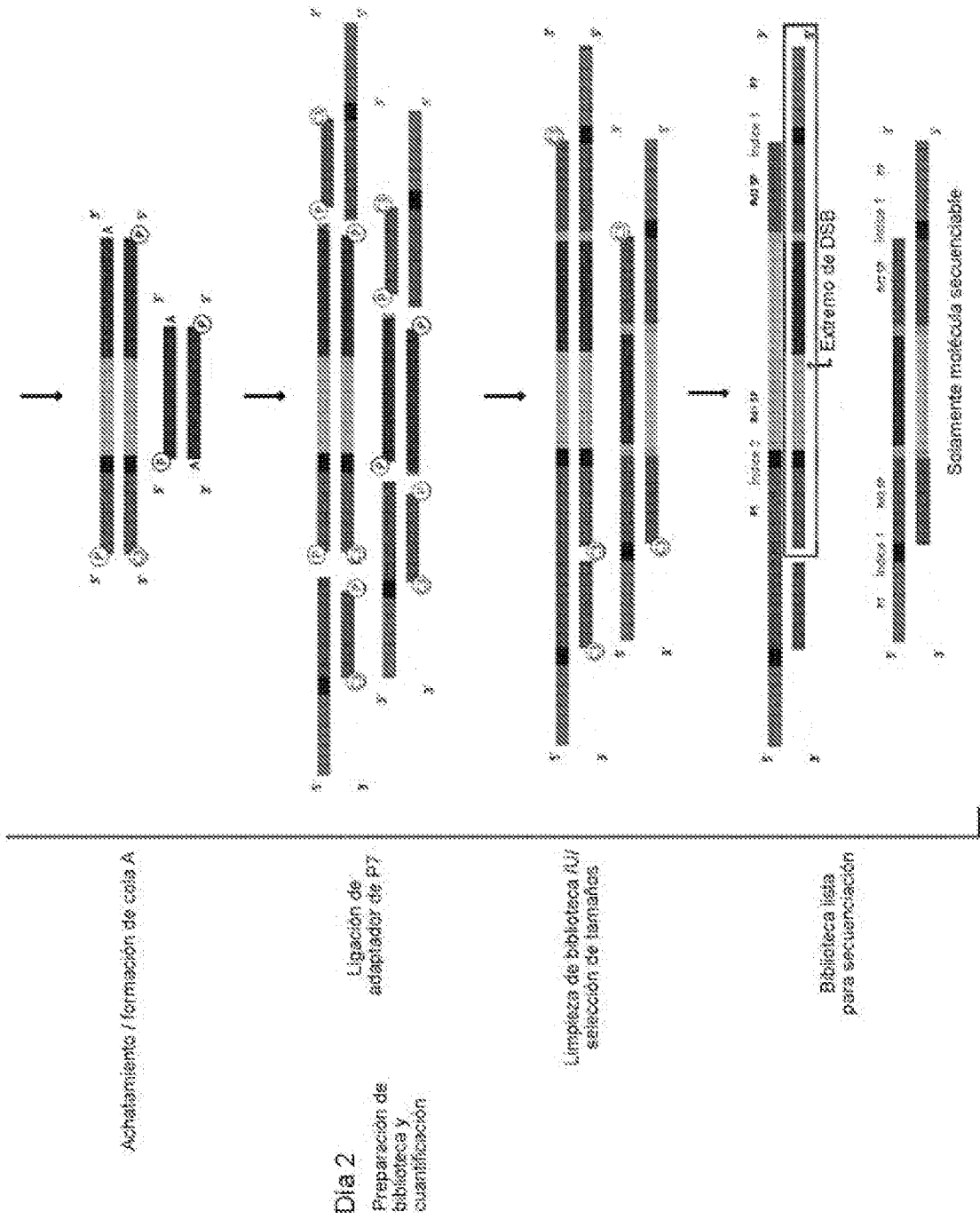
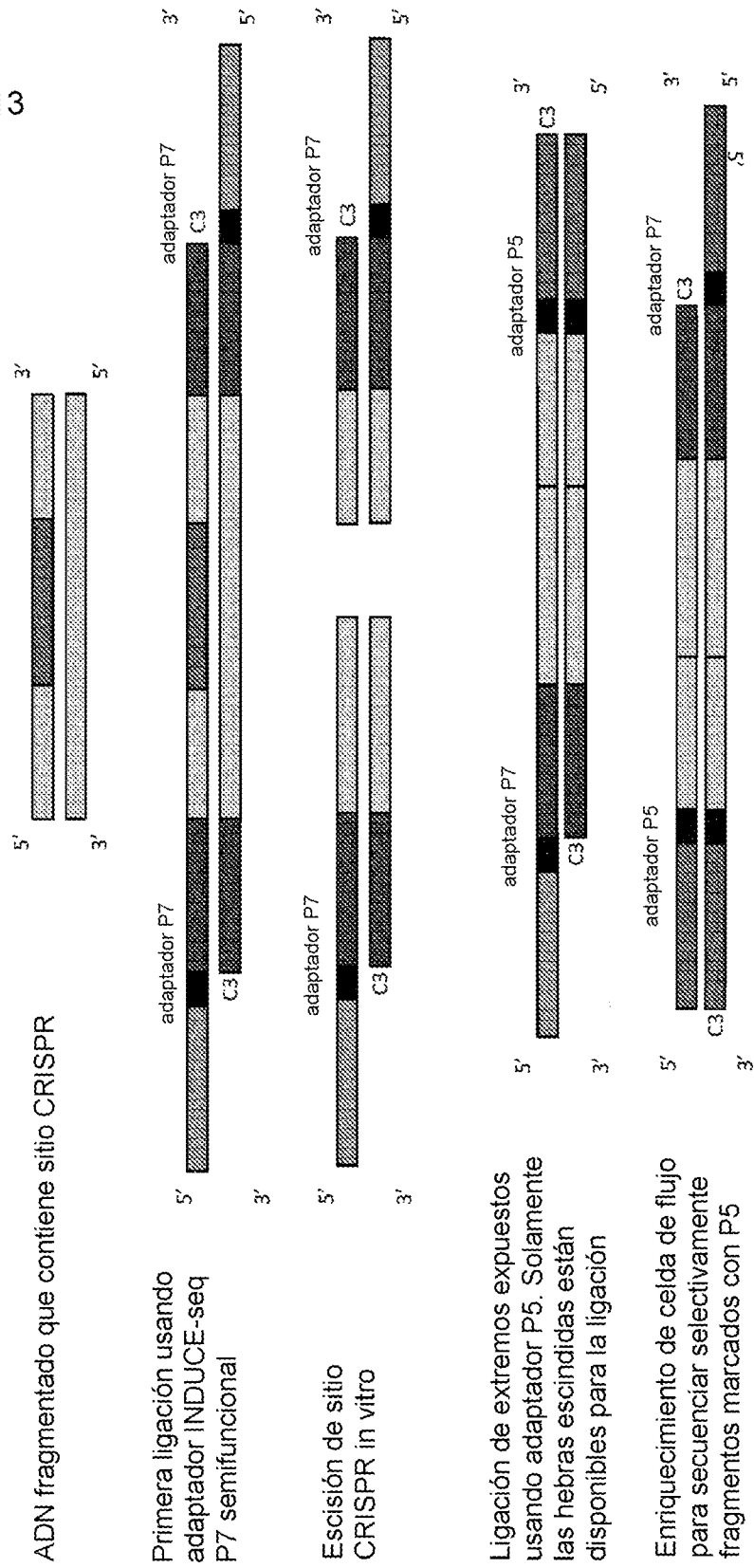


Figura 12 (cont.)

Figura 13



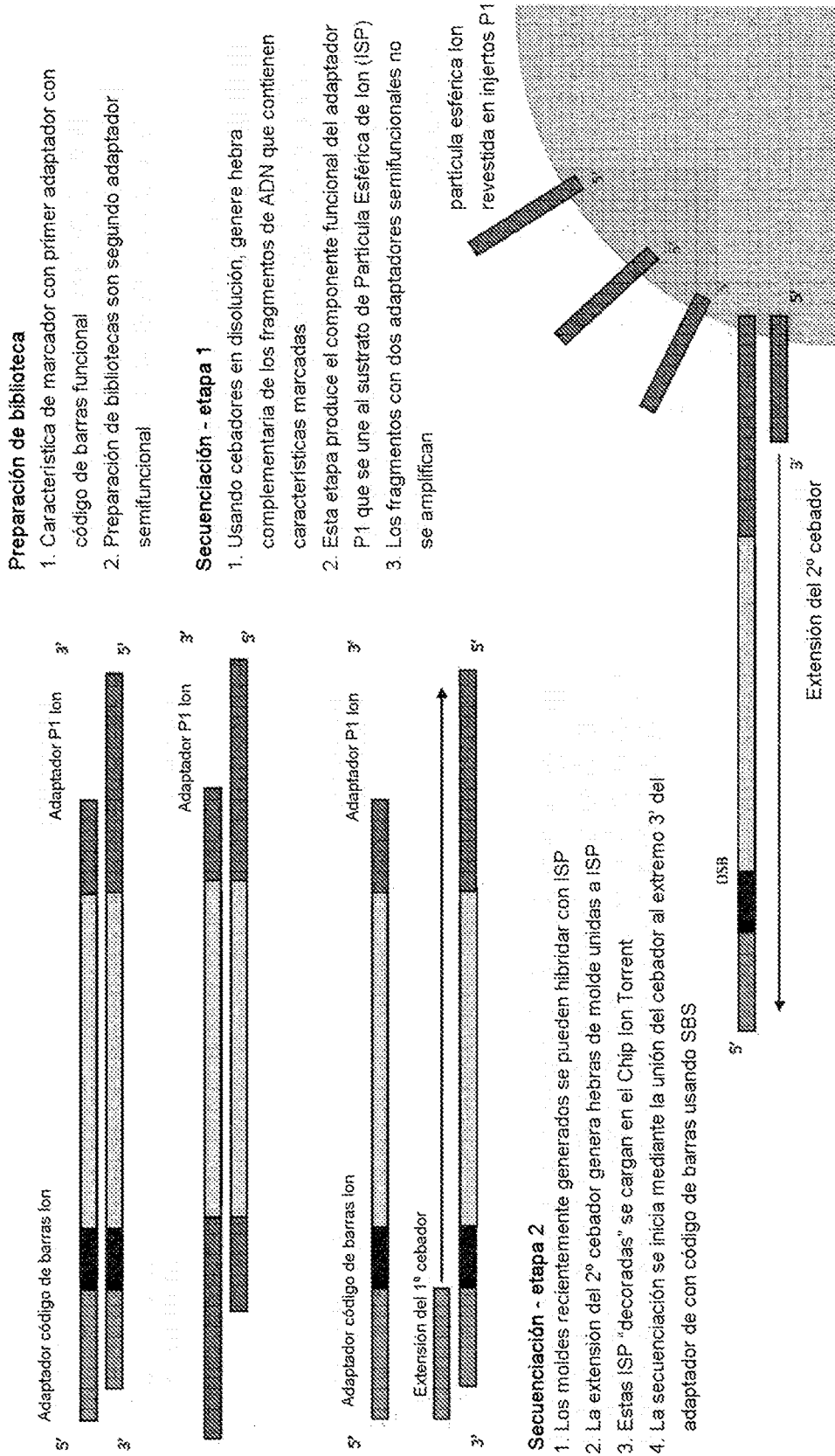


Figura 14

Adaptadores Torrent INDUCE-seq

Ion Xpress™ código de barras 1 (orientación 5'-3' para código de barras en **negrita subrayado**) incluyendo Cola T

Cebador 1 CCAATCTGATGCTT-**g-gggtttttcc** (SEQ ID NO: 106)
 5' CGAGTCAATGCTT-g-gggtttttccGCACCTAGG**ST**AACGAT***T**-NNNNNNNNNNNNNNNNNN-A 3' (SEQ ID NO: 107)
 3' /3SpC3/ GGTAGATTAGGGA***g-gggtttttcc**CTGAGTCA**TT**CGCTA~**P**A~NNNNNNNNNNNNNNNNNN~**F** 5' (SEQ ID NO: 108)
 Cebador 2
 Igual que para injerto ISP ^A Posición de DSE

Adaptador P1 Ion semifuncional (incluyendo cola T)

| | | |
|----------------------------------------------------------|-------------------|---------------------|
| F -ATCACCGACTGCCCATAGAGA/3SpC3/ | Secuencia perdida | 3' (SEQ ID NO: 109) |
| T *TAGTGGCTGACGGGTAATCTCTGCTTTTGGGCTGCGCATCAGCGAA | | 5' (SEQ ID NO: 110) |
| CGTTTGGGCTGGGCAACAGCGAA | | (SEQ ID NO: 111) |

Figura 15

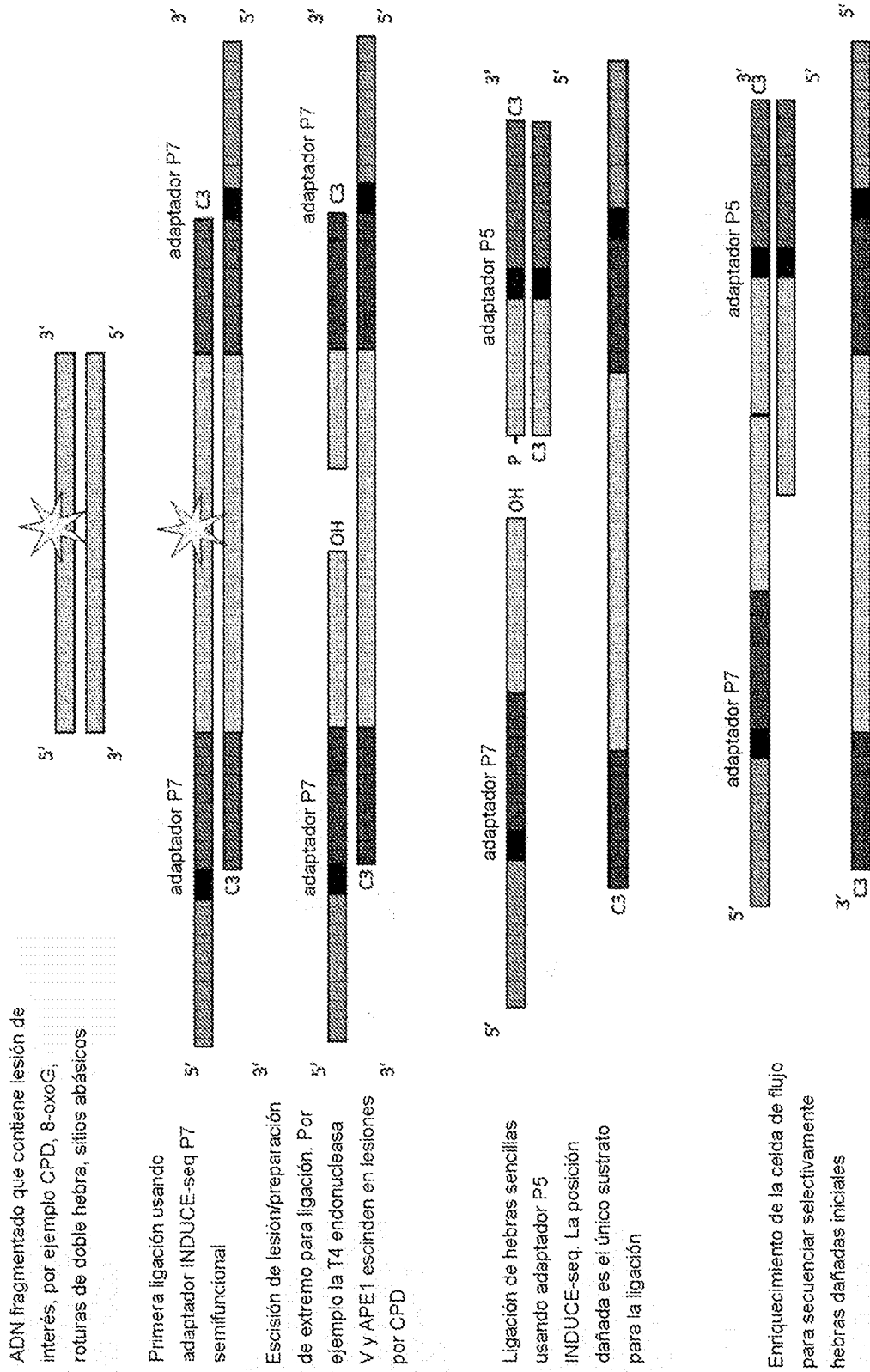


Figura 16

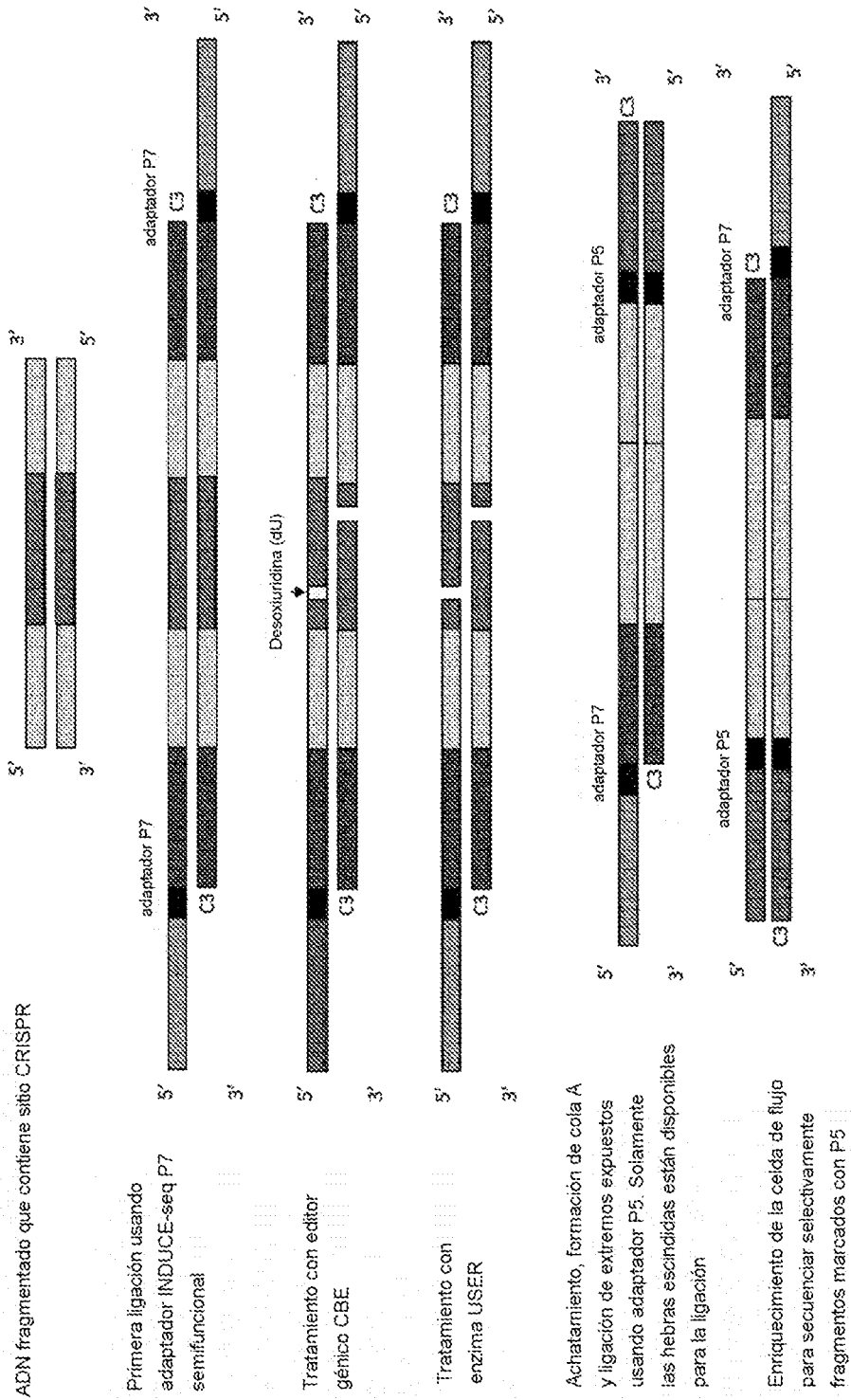


Figura 17

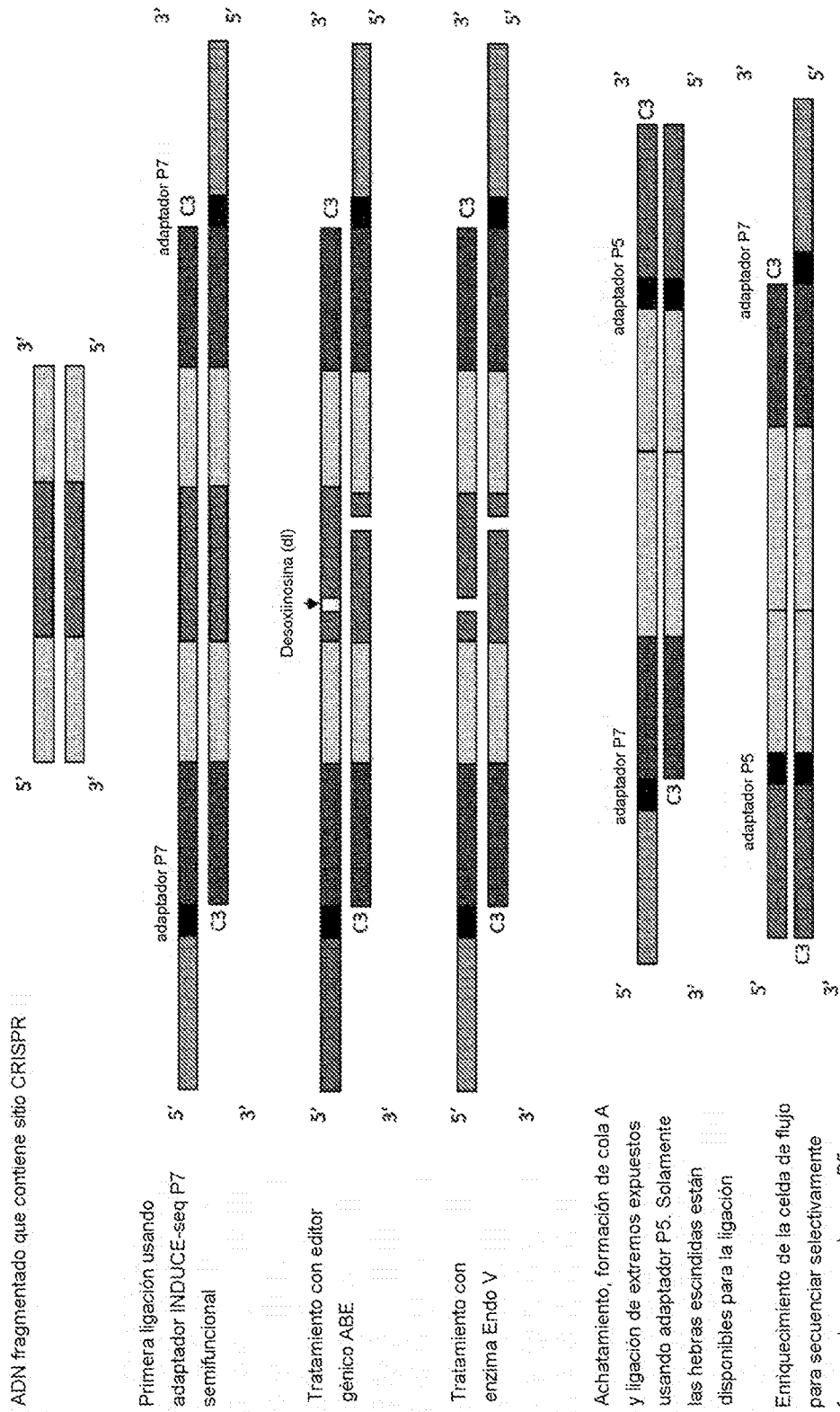


Figura 18