

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 730 378**

51 Int. Cl.:

C12N 15/11 (2006.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **26.08.2015 PCT/US2015/047046**
- 87 Fecha y número de publicación internacional: **03.03.2016 WO16033246**
- 96 Fecha de presentación y número de la solicitud europea: **26.08.2015 E 15763698 (6)**
- 97 Fecha y número de publicación de la concesión europea: **20.03.2019 EP 3186376**

54 Título: **Procedimientos para incrementar la eficiencia de la modificación mediada por Cas9**

30 Prioridad:

27.08.2014 US 201462042358 P
08.09.2014 US 201462047495 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.11.2019

73 Titular/es:

CARIBOU BIOSCIENCES, INC. (100.0%)
2929 7th Street, Suite 105
Berkeley, CA 94710, US

72 Inventor/es:

CAMERON, PETER, SEAN;
HAURWITZ, RACHEL, E.;
MAY, ANDREW, P.;
NYE, CHRISTOPHER, H y
VAN OVERBEEK, MEGAN

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 730 378 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

Procedimientos para incrementar la eficiencia de la modificación mediada por Cas9

Campo técnico

5 La presente invención se refiere a los sistemas CRISPR-Cas9 de Tipo II para uso en el aumento de la eficiencia de la modificación del genoma mediada por Cas9 ya sea disminuyendo el número de rompimientos de cadenas dobles no deseadas del ácido nucleico, y/o mejorando la reparación dirigida por homología de un ácido nucleico objetivo escindido.

Antecedentes de la invención

10 Las repeticiones palindrómicas cortas agrupadas en intervalos regulares (CRISPR) y las proteínas Cas9 asociadas constituyen el sistema CRISPR-Cas9. Este sistema proporciona inmunidad adaptativa contra el ADN foráneo en bacterias (Barintervalou, R., et al., "CRISPR provides acquired resistance against viruses in prokaryotes", *Science* (2007) 315: 1709-1712 ; Makarova, K. S., et al., "Evolution and classification of the CRISPR-Cas systems", *Nat Rev Microbiol* (2011) 9: 467-477; Garneau, J. E., et al., "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA", *Nature* (2010) 468: 67-71; Sapranaukas, R., et al., "The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*", *Nucleic Acids Res* (2011) 39: 9275-9282).

15 La endonucleasa Cas9 guiada por ARN se dirige específicamente a y escinde el ADN de una manera dependiente de la secuencia (Gasiunas, G., et al., "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria", *Proc Natl Acad Sci USA* (2012) 109: E2579-E2586; Jinek, M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* (2012) 337: 816-821; Sternberg, S. H., et al., "DNA interrogation by the CRISPR RNA-guided endonuclease Cas9", *Nature* (2014) 507: 62; Deltcheva, E., et al., "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III", *Nature* (2011) 471: 602-607), y se ha utilizado ampliamente para la edición programable del genoma en una variedad de organismos y sistemas modelo (Cong, L., et al., "Multiplex genome engineering using CRISPR/Cas systems", *Science* (2013) 339: 819-823; Jiang, W., et al., "RNA-guided editing of bacterial genomes using CRISPR-Cas systems", *Nat. Biotechnol.* (2013) 31: 233-239; Sander, J. D. & Joung, J. K., "CRISPR/Cas systems for editing, regulating and targeting genomes", *Nature Biotechnol.* (2014) 32: 347-355).

20 Jinek, M., et al., ("A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* (2012) 337: 816-21) mostraron que en un subconjunto de sistemas asociados a CRISPR (Cas), el ARNcr maduro que está apareado por bases para activar ARNcr en forma trans (ARNcrtra) forma una estructura de ARN de dos partes, también llamada "guía doble", que dirige la proteína Cas9 asociada con CRISPR para introducir rompimientos de cadena doble en el ADN diana. En sitios complementarios a la secuencia guía de ARNcr (espaciador), el dominio de nucleasa HNH de Cas9 escinde la cadena complementaria y el dominio tipo RuvC de Cas9 escinde la cadena no complementaria. Las moléculas de doble ARNcr/ARNcrtra se diseñaron en moléculas de ARNcr/ARNcrtra de cadena sencilla. Estas ARNcr/ARNcrtra de cadena sencilla dirigieron la escisión del ADN de cadena doble de Cas9 específica de la secuencia objetivo.

25 Sin embargo, las nucleasas específicas del sitio tales como Cas9 pueden introducir rompimientos de cadena doble en el ADN en sitios no pretendidos y/o ubicaciones incorrectas, denominados "efectos alejados del objetivo". En consecuencia, los procedimientos para reducir o eliminar los rompimientos de ADN alejados del objetivo son muy deseables.

30 Además, los rompimientos de cadena doble del ADN se pueden reparar, por ejemplo, mediante uniones de extremos no homólogos (NHEJ) o reparación dirigida por homología (HDR). La reparación fiel por HDR es ineficiente en los rompimientos dirigidos al sitio del ácido nucleico objetivo debido a que otros mecanismos celulares pueden resultar en la incorporación de ácidos nucleicos en el sitio de un rompimiento de cadena doble o un corte de una sola cadena. Es evidente que existe una clara necesidad de desarrollar estrategias novedosas que mitiguen o eliminen los eventos alejados del objetivo de edición del genoma y aumenten la eficacia de la inserción de nuevo material en los sitios cortados por nucleasas dirigidas al sitio tales como Cas9.

Sumario

35 La invención se define por las reivindicaciones adjuntas. En un aspecto, la divulgación proporciona un procedimiento para reducir la escisión de la nucleasa no deseada que comprende: poner en contacto un primer complejo que comprende una Cas9 catalíticamente activa y un ARN guía con un ácido nucleico objetivo; poner en contacto un segundo complejo que comprende una Cas9 catalíticamente inactiva (dCas9) y un ARN guía con un ácido nucleico alejado del objetivo; y escindir el ácido nucleico objetivo con el primer complejo, en el que el segundo complejo evita que el primer complejo escinda el ácido nucleico alejado del objetivo. En algunos aspectos, la Cas9 activa comprende al menos un 25% de identidad de aminoácidos con los motivos de sitios activos de HNH y RuvC de una Cas9 de *Streptococcus pyogenes*, tal como al menos el 50%, 75%, 95%, 99% y la identidad completa de aminoácidos, o cualquier porcentaje entre 25% y 100%, con una Cas9 de *S. pyogenes*.

- En algunos aspectos, la Cas9 activa comprende al menos un 25% de identidad de aminoácidos con los motivos del sitio activo de HNH y RuvC de una Cas9 de *Streptococcus thermophilus*, tal como al menos el 50%, 75%, 95%, 99% e identidad completa de aminoácidos, o cualquier porcentaje entre el 25% y el 100%, con una Cas9 de *S. thermophilus*.
- 5 En algunos aspectos, la Cas9 activa comprende al menos un 25% de identidad de aminoácidos con los motivos del sitio activo de HNH y RuvC de una Cas9 de *Staphylococcus aureus*, tal como al menos el 50%, 75%, 95%, 99% y una identidad completa de aminoácidos, o cualquier porcentaje entre el 25% y el 100%, con una Cas9 de *S. aureus*. En algunas realizaciones, la Cas9 activa comprende al menos un 25% de identidad de aminoácidos con los motivos del sitio activo de HNH y RuvC de una Cas9 de *Neisseria meningitidis*, tal como al menos el 50%, 75%, 95%, 99% y una identidad completa de aminoácidos, o cualquier porcentaje entre el 25% y el 100%, con una Cas9 de *N. meningitidis*.
- 10 En algunos aspectos, la Cas9 catalíticamente inactiva comprende una mutación en uno o ambos de sus dominios de nucleasa. En algunos aspectos, la dCas9 es al menos 80% catalíticamente inactiva en comparación con una Cas9 de tipo silvestre.
- En algunos aspectos, el primer complejo es capaz de unirse al ácido nucleico alejado del objetivo. En algunos aspectos, la unión y/o escisión del primer complejo al ácido nucleico alejado del objetivo se reduce en al menos un 30%. En algunos aspectos, la unión del primer complejo al ácido nucleico alejado del objetivo se reduce en al menos un 70%.
- 15 En algunos aspectos, la escisión comprende la introducción de un rompimiento de cadena doble. En algunos aspectos, la división comprende la introducción de un rompimiento de cadena sencilla. En algunas realizaciones, el ácido nucleico objetivo es el ADN. En algunos aspectos, el ácido nucleico objetivo es un ADN de cadena doble.
- 20 En otro aspecto, la divulgación proporciona una composición que comprende: dos polipéptidos dirigidos al sitio de Cas9, en el que los dos polipéptidos dirigidos al sitio comprenden una mutación en uno de sus dominios de nucleasa, en el que los dos polipéptidos dirigidos al sitio están configurados para unirse y escinden la misma cadena de un ácido nucleico objetivo de cadena doble.
- 25 En algunos aspectos, los dos polipéptidos dirigidos al sitio comprenden al menos un 10% de identidad de aminoácidos con un dominio de nucleasa de Cas9 de *S. pyogenes*.
- En algunos aspectos, la mutación comprende una mutación D10A. En algunos aspectos, la mutación comprende una mutación H840A. En algunos aspectos, el ácido nucleico objetivo es ADN.
- 30 En algunos aspectos, los dos polipéptidos dirigidos al sitio están unidos a la cadena sentido del ácido nucleico objetivo de cadena doble. En algunos aspectos, los dos polipéptidos dirigidos al sitio están unidos a la cadena antisentido del ácido nucleico objetivo de cadena doble. En algunos aspectos, la composición comprende además un polinucleótido donante. En algunos aspectos, el polinucleótido donante es de cadena sencilla. En algunos aspectos, el polinucleótido donante es de cadena doble. En algunos aspectos, el polinucleótido donante es parcialmente de cadena sencilla y parcialmente de cadena doble.
- 35 En una realización, se proporciona un procedimiento para reducir la unión y/o la escisión de un ácido nucleico alejado del objetivo por un complejo que comprende una proteína Cas9 catalíticamente activa y un polinucleótido guía. El procedimiento comprende: (a) poner en contacto un primer complejo con un ácido nucleico objetivo seleccionado, en el que dicho primer complejo comprende: (i) una proteína Cas9 catalíticamente activa y (ii) un primer polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a dicho ácido nucleico objetivo seleccionado; y (b) poner en contacto un segundo complejo con un ácido nucleico alejado del objetivo, en el que dicho segundo complejo comprende (i) una proteína Cas9 catalíticamente inactiva (proteína dCas9) que no escinde el ácido nucleico alejado del objetivo y (ii) un segundo polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a dicho ácido nucleico alejado del objetivo, reduciendo así la unión y/o la escisión por dicho primer complejo de dicho ácido nucleico alejado del objetivo, en el que el ácido nucleico objetivo seleccionado es ADN, y en el que el procedimiento no comprende el tratamiento del cuerpo humano o animal mediante cirugía o terapia practicada en el cuerpo humano o animal.
- 40 En algunas realizaciones, la proteína Cas9 catalíticamente activa comprende al menos un 75% de identidad de aminoácidos con una Cas9 de *S. pyogenes*, con la condición de que la proteína Cas9 retenga la actividad catalítica. En ciertas realizaciones, la proteína Cas9 catalíticamente activa comprende al menos un 95% de identidad de aminoácidos con una Cas9 de *S. pyogenes*, con la condición de que la proteína Cas9 retenga la actividad catalítica. En realizaciones adicionales del procedimiento, la proteína Cas9 catalíticamente activa es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga.
- 45 En realizaciones adicionales, la proteína dCas9 comprende al menos una mutación en uno o más dominios de endonucleasa para volver la proteína dCas9 catalíticamente inactiva. En algunas realizaciones, la proteína dCas9 comprende al menos un 75% de identidad de aminoácidos con una proteína Cas9 de *S. pyogenes*. En otras realizaciones, la proteína dCas9 comprende al menos un 75% de identidad de aminoácidos con una proteína Cas9 de *S. pyogenes*. En realizaciones adicionales, la proteína dCas9 es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga con al menos una mutación en uno o más dominios de endonucleasa para volver la proteína
- 55

5 Cas9 ortóloga catalíticamente inactiva. En ciertas realizaciones, las una o más mutaciones están en un dominio RuvC-1, tal como una mutación D10A, numerada con respecto a Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga. En otras realizaciones, las una o más mutaciones están en el dominio HNH, tal como una mutación H840A, numerada con relación a Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga. En realizaciones adicionales, las una o más mutaciones comprenden una mutación D10A y una mutación H840A, numeradas con respecto a Cas9 de *S. pyogenes*, o las mutaciones correspondientes en una proteína Cas9 ortóloga.

En la invención, el ácido nucleico objetivo seleccionado es ADN, tal como ADN de cadena doble.

10 En realizaciones adicionales, el ácido nucleico objetivo seleccionado se escinde para proporcionar un sitio de escisión y el procedimiento comprende además modificar el ácido nucleico objetivo, tal como insertando al menos una parte del polinucleótido donante en el sitio de escisión. En otras realizaciones, la modificación comprende eliminar uno o más nucleótidos en el sitio de escisión.

En realizaciones adicionales, el procedimiento se realiza en una célula, tal como una célula eucariota, o *in vitro*.

15 En otra divulgación, se proporciona un procedimiento para modificar un ácido nucleico objetivo que comprende: poner en contacto dos complejos con la misma cadena del ácido nucleico objetivo, en el que cada uno de los dos complejos comprende un polipéptido dirigido al sitio y un ácido nucleico dirigido al ácido nucleico, en el que los dos polipéptidos dirigidos al sitio comprenden una mutación en uno de sus dominios de nucleasa; y modificando el ácido nucleico objetivo. En algunos aspectos, el ácido nucleico dirigido a un ácido nucleico de uno de los dos complejos se dirige a una ubicación diferente en el ácido nucleico objetivo que el ácido nucleico dirigido a un ácido nucleico del otro de los dos complejos.

20 En algunos aspectos, los dos polipéptidos dirigidos al sitio comprenden al menos un 75% de identidad de aminoácidos con Cas9 de *S. pyogenes*. En algunos aspectos, los dos polipéptidos dirigidos al sitio comprenden al menos un 10% de identidad de aminoácidos con un dominio de nucleasa de Cas9 de *S. pyogenes*. En algunos aspectos, la mutación comprende una mutación D10A. En algunos aspectos, la mutación comprende una mutación H840A. En algunos aspectos, el ácido nucleico objetivo es ADN.

30 En algunos aspectos, los dos polipéptidos dirigidos al sitio están unidos a la cadena sentido del ácido nucleico objetivo de cadena doble. En algunos aspectos, los dos polipéptidos dirigidos al sitio están unidos a la cadena antisentido del ácido nucleico objetivo de cadena doble. En algunos aspectos, la modificación comprende dividir la misma cadena del ácido nucleico objetivo. En algunas realizaciones, la escisión comprende un rompimiento de cadena sencilla. En algunos aspectos, el procedimiento comprende además insertar un polinucleótido donante en el ácido nucleico objetivo. En algunos aspectos, el polinucleótido donante es de cadena sencilla. En algunos aspectos, el polinucleótido donante es de cadena doble. En algunos aspectos, el polinucleótido donante es parcialmente de cadena sencilla y parcialmente de cadena doble.

35 En otra divulgación, la invención está dirigida a un procedimiento para escindir una sola cadena de un ácido nucleico objetivo que comprende poner en contacto el primero y segundo complejos en ubicaciones separadas en la misma cadena de una molécula de ácido nucleico. El primer complejo comprende (i) una primera proteína Cas9 con una mutación en un dominio de endonucleasa de la misma para hacer que la proteína Cas9 sea una nickasa; y (ii) un primer polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a un primer ácido nucleico objetivo. El segundo complejo comprende (i) una segunda proteína Cas9 con una mutación en un dominio de endonucleasa de la misma, para hacer que la proteína Cas9 sea una nickasa; y (ii) un segundo polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a un segundo ácido nucleico objetivo; en el que la primera y la segunda proteínas Cas9 escinden una sola cadena de dicha molécula de ácido nucleico en sitios apartados en la misma cadena, para producir un rompimiento de una sola cadena.

45 En algunos aspectos, la primera proteína Cas9 y/o la segunda proteína Cas9 comprenden al menos un 75% de identidad de aminoácidos con una Cas9 de *S. pyogenes*. En ciertos aspectos, la proteína Cas9 comprende al menos un 95% de identidad de aminoácidos con una Cas9 de *S. pyogenes*. En aspectos adicionales del procedimiento, la primera proteína Cas9 y/o la segunda proteína Cas9 es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga con una mutación en un dominio de endonucleasa de la misma, para hacer que la proteína Cas9 ortóloga sea una nickasa. En ciertos aspectos, las una o más mutaciones se encuentran en un dominio RuvC-1, tal como una mutación D10A, numerada con respecto a Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga. En otros aspectos, las una o más mutaciones se encuentran en el dominio HNH, tal como una mutación H840A, numerada en relación con Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga.

55 En aspectos adicionales, el ácido nucleico objetivo es un ADN de cadena doble y los complejos se unen y escinden la cadena antisentido del ADN de cadena doble. En otros aspectos, el ácido nucleico objetivo es un ADN de cadena doble y los complejos se unen y escinden la cadena sentido del ADN de cadena doble.

En aspectos adicionales, el procedimiento comprende además modificar el ácido nucleico objetivo, tal como mediante la inserción de al menos una porción del polinucleótido donante en el ácido nucleico objetivo en el

rompimiento de una sola cadena. En ciertos aspectos, el polinucleótido donante es de cadena sencilla. En aspectos adicionales, la inserción se realiza utilizando la reparación dirigida por homología del polinucleótido donante con el ácido nucleico objetivo.

En aspectos adicionales, el procedimiento se realiza en una célula, tal como una célula eucariótica, o *in vitro*.

- 5 En otras divulgaciones adicionales, se proporciona un procedimiento para la reparación dirigida por homología (HDR) en un ácido nucleico objetivo. El procedimiento comprende: (a) poner en contacto un primer complejo con un primer ácido nucleico objetivo, en el que dicho primer complejo comprende: (i) una proteína Cas9 catalíticamente activa y (ii) un primer polinucleótido guía, tal como un ARN_g, que comprende un espaciador adaptado para unirse a dicho primer ácido nucleico objetivo, en el que dicho primer complejo escinde el primer ácido nucleico objetivo; y (b)
- 10 poner en contacto un segundo complejo con un segundo ácido nucleico objetivo, en el que dicho segundo complejo comprende: (i) una primera proteína Cas9 catalíticamente inactiva (proteína dCas9) que comprende al menos una mutación en uno o más dominios de endonucleasa para volver la proteína dCas9 catalíticamente inactiva de tal forma que la proteína dCas9 no escinda al segundo ácido nucleico objetivo, y (ii) un segundo polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a dicho segundo ácido nucleico objetivo, en el
- 15 que el segundo complejo comprende uno extremo de un polinucleótido donante asociado con el mismo y configurado cerca del primer ácido nucleico objetivo escindido; en el que al menos una porción del polinucleótido donante se inserta en el primer ácido nucleico objetivo a través de HDR.

En ciertos aspectos, el segundo ácido nucleico objetivo está secuencia arriba del primer ácido nucleico objetivo. En otros aspectos, el segundo ácido nucleico objetivo está secuencia abajo del primer ácido nucleico objetivo.

- 20 En ciertos aspectos del procedimiento anterior, el extremo 5' del polinucleótido donante está asociado con el segundo complejo. En otros aspectos, el extremo 3' del polinucleótido donante está asociado con el segundo complejo.

- En aspectos adicionales, el procedimiento comprende además: (c) poner en contacto un tercer complejo con un tercer ácido nucleico objetivo, en el que el tercer ácido nucleico objetivo se coloca secuencia abajo del primer ácido nucleico objetivo cuando el primer ácido nucleico objetivo está secuencia abajo del segundo ácido nucleico objetivo, o en el que el tercer ácido nucleico objetivo está posicionado secuencia arriba del primer ácido nucleico objetivo cuando el primer ácido nucleico objetivo está secuencia arriba del segundo ácido nucleico objetivo, en el que dicho tercer complejo comprende: (i) una segunda proteína dCas9 que comprende al menos una mutación en uno o más dominios de endonucleasa para volver la segunda proteína dCas9 catalíticamente inactiva de manera que la
- 25 segunda proteína dCas9 no escinda el tercer ácido nucleico objetivo, y (ii) un tercer polinucleótido guía, tal como ARN_g, que comprende un espaciador adaptado para unirse a dicho tercer ácido nucleico objetivo, y en el que el tercer complejo comprende el otro extremo del polinucleótido donante asociado con el segundo complejo. En ciertos aspectos, el extremo 5' del polinucleótido donante está asociado con el segundo complejo y el extremo 3' del polinucleótido donante está asociado con el tercer complejo. En otros aspectos, el extremo 3' del polinucleótido donante está asociado con el segundo complejo y el extremo 5' del polinucleótido donante está asociado con el
- 30 tercer complejo. En otros aspectos, el extremo 3' del polinucleótido donante está asociado con el tercer complejo.

- En otros aspectos, la proteína Cas9 comprende al menos un 75% de identidad de aminoácidos con una Cas9 de *S. pyogenes*, con la condición de que la proteína Cas9 retenga la actividad catalítica. En ciertos aspectos, la proteína Cas9 comprende al menos un 95% de identidad de aminoácidos con una Cas9 de *S. pyogenes*, con la condición de que la proteína Cas9 retenga la actividad catalítica. En aspectos adicionales del procedimiento, la proteína Cas9 es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga.
- 35

- En aspectos adicionales, la proteína dCas9 comprende al menos un 75% de identidad de aminoácidos con una proteína Cas9 de *S. pyogenes*. En otros aspectos, la proteína dCas9 comprende al menos un 75% de identidad de aminoácidos con una proteína Cas9 de *S. pyogenes*. En aspectos adicionales, la proteína dCas9 es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga con al menos una mutación en uno o más dominios de endonucleasa para volver la proteína Cas9 ortóloga catalíticamente inactiva. En ciertos aspectos, las una o más mutaciones se encuentran en un dominio RuvC-1, tal como una mutación D10A, numerada con respecto a Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga. En otros aspectos, las una o más mutaciones se encuentran en el dominio HNH, tal como una mutación H840A, numerada con respecto a Cas9 de *S. pyogenes*, o la mutación correspondiente en una proteína Cas9 ortóloga. En aspectos adicionales, las una o más mutaciones comprenden una mutación D10A y una mutación H840A, numeradas con respecto a Cas9 de *S. pyogenes*, o las mutaciones correspondientes en una proteína Cas9 ortóloga.
- 40
- 45
- 50

En aspectos adicionales, el ácido nucleico objetivo seleccionado es ADN, tal como ADN de cadena doble.

En otros aspectos, el procedimiento se realiza en una célula, tal como una célula eucariota, o *in vitro*.

- 55 Estos aspectos y otras realizaciones de los procedimientos para aumentar la eficiencia de la ingeniería mediada por Cas9 y/o la reparación de HDR se les ocurrirán fácilmente a los expertos en la técnica en vista de la divulgación de este documento.

Breve descripción de los dibujos

Las figuras que no pertenecen a la invención son solo para fines ilustrativos. La Fig. 1A y la Fig. 1B presenta ejemplos ilustrativos de los ARN asociados a CRISPR-Cas9 de Tipo II. La Fig. 1A muestra un componente de dos RNA Tipo II CRISPR-Cas9 que comprende un ARNcr (Fig. 1A, 101) y un ARNcrtra (Fig. 1A, 102), también conocido como ARN guía dual. La Fig. 1B ilustra la formación de enlaces de hidrógeno de pares de bases entre el ARNcr y el ARNcrtra para formar una estructura secundaria (véase la solicitud de patente publicada de Estados Unidos No. 2014-0068797, publicada el 6 de marzo de 2014; véase también Jinek M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* (2012) 337: 816-821). Las figuras presentan una visión general y una nomenclatura de los elementos estructurales secundarios del ARNcr y ARNcrtra de la Cas9 de *S. pyogenes* que incluye lo siguiente: un elemento espaciador (Fig. 1B, 103); un primer elemento del tallo que comprende un elemento inferior del tallo (Fig. 1B, 104), un elemento de protuberancia que comprende nucleótidos no emparejados (Fig. 1B, 105) y un elemento superior del tallo (Fig. 1B, 106); un elemento de vínculo (Fig. 1B, 107); un segundo elemento de horquilla que comprende un segundo elemento del tallo (Fig. 1B, 108); y un tercer elemento de horquilla que comprende un tercer elemento del tallo (Fig. 1B, 109). Las figuras no se presentan proporcionalmente ni son a escala. Las ubicaciones de los indicadores son aproximadas.

La Fig. 2 muestra otro ejemplo de un ARN asociado a CRISPR-Cas9 tipo II. La figura ilustra un ARN guía único (ARNgu) en el que el ARNcr se une covalentemente al ARNcrtra y forma una estructura secundaria polinucleotídica de ARN a través del enlace de hidrógeno del par de bases (véase, por ejemplo, la solicitud de patente publicada de los Estados Unidos No. 2014-0068797, publicada el 6 de marzo del 2014). La figura presenta una descripción general y nomenclatura de los elementos estructurales secundarios de un ARNgu de la Cas9 de *S. pyogenes* que incluye lo siguiente: un elemento espaciador (Fig. 2, 201); un primer elemento del tallo que comprende un elemento inferior del tallo (Fig. 2, 202), un elemento de protuberancia que comprende nucleótidos no apareados (Fig. 2, 205) y un elemento superior del tallo (Fig. 2, 203); un elemento de bucle (Fig. 2, 204) que comprende nucleótidos no apareados; (un primer elemento de horquilla comprende el primer elemento del tallo y el elemento del bucle); un elemento de vínculo (Fig. 2, 206); un segundo elemento de horquilla que comprende un segundo elemento del tallo (Fig. 2, 207); y un tercer elemento de horquilla que comprende un tercer elemento del tallo (Fig. 2, 208). (Véase, por ejemplo, las Figuras 1 y 3 de Briner, AE, et al., "Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality", *Molecular Cell* (2014) 56: 333-339.) La figura no se representa proporcionalmente ni a escala. Las ubicaciones de los indicadores son aproximadas.

La Fig. 3A y la Fig. 3B se relacionan con información estructural para un complejo de ARNgu/proteína Cas y una proteína Cas, respectivamente. La Fig. 3A proporciona un modelo basado en la estructura cristalina de Cas9 de *S. pyogenes* (SpyCas9) en un complejo activo con ARNgu (Anders C., et al., "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease," *Nature* (2014) 513: 569-573). Los estudios estructurales de SpyCas9 mostraron que la proteína presenta una arquitectura bilobulada que comprende el lóbulo de nucleasa catalítica y el lóbulo α -helicoidal de la enzima (véase, Jinek M., et al., "Structures of Cas9 endonucleases reveal RNA-mediated conformational activation", *Science* (2014) 343: 1247997; Anders C., et al., "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease", *Nature* (2014) 513: 569-573). En la Fig. 3A, el lóbulo α -helicoidal (Fig. 3A, dominio helicoidal) se muestra como el lóbulo más oscuro; el lóbulo de la nucleasa catalítica (Fig. 3A, lóbulo de la nucleasa catalítica) se muestra en gris claro y la cadena principal de ARNgu se muestra en negro (Fig. 3A, ARNgu). Se indica la ubicación relativa del extremo 3' del ARNgu (Fig. 3A, ARNgu del extremo 3'). El ARN espaciador del ARNgu no es visible porque está rodeado por los dos lóbulos de la proteína. Se indica la ubicación relativa del extremo 5' del ARNgu (Fig. 3A, ARNgu del extremo 5') y el ARN espaciador del ARNgu está ubicado en la región del extremo 5' del ARNgu. Un residuo de cisteína (Fig. 3A, Cys de SpyCas9 de tipo silvestre) en SpyCas9 de tipo silvestre se identifica en la presente divulgación como un sitio de entrecruzamiento disponible. En la Fig. 3A, el lóbulo de la nucleasa catalítica se muestra como el lóbulo más claro en el que se indican las posiciones relativas de los dominios RuvC (Fig. 3A, RuvC; dominio homólogo de RNasa H) y de nucleasa HNH (Fig. 3A, HNH; dominio homólogo de nucleasa HNH). Los dominios de nucleasa RuvC y HNH, cuando están activos, cortan cada uno una cadena de ADN diferente en el ADN diana. El dominio del extremo terminal C (Fig. 3A, CTD) está involucrado en el reconocimiento de los motivos adyacentes protoespaciadores (PAM) en el ADN diana. La Fig. 3B presenta un modelo de la disposición del dominio de SpyCas9 en relación con su estructura de secuencia primaria. En la Fig. 3B, tres regiones de la secuencia primaria corresponden al dominio RuvC (Fig. 3B, RuvC-I (aminoácidos 1-78), RuvC-II (aminoácidos 719-765) y RuvC-III (aminoácidos 926-1102)). Una región corresponde al dominio helicoidal (Fig. 3B, dominio helicoidal (aminoácidos 79-718)). Una región corresponde al dominio HNH (Fig. 3B, HNH (aminoácidos 766-925)). Una región corresponde al dominio CTD (Fig. 3B, CTD (aminoácidos 1103-1368)). En la Fig. 3B, las regiones de la secuencia primaria correspondiente al lóbulo α -helicoidal (Fig. 3B, lóbulo alfa-helicoidal) y el lóbulo del dominio de la nucleasa (Fig. 3B, lóbulo del dominio de la nucleasa) se indican con corchetes.

La Fig. 4 representa un ejemplo de realización de la unión no deseada y la escisión durante la modificación del genoma. En esta realización, un ácido nucleico objetivo (Fig. 4, 115) se pone en contacto con un complejo que comprende un polipéptido dirigido al sitio (por ejemplo, Cas9) (Fig. 4, 105) y un ácido nucleico dirigido al ácido nucleico (por ejemplo, ARNgu o ARN guía dual) (Fig. 4, 110). El complejo que comprende la Cas9 se une a un ácido nucleico objetivo (Fig. 4, 120). En algunos casos, el complejo se une a un ácido nucleico alejado del objetivo (Fig. 4, 125). En una etapa de escisión (Fig. 4, 130), la Cas9 del complejo puede escindir el ácido nucleico objetivo (Fig. 4, 120) y el ácido nucleico alejado del objetivo, generando así efectos alejados del objetivo.

La Fig. 5 representa una realización ejemplar de un procedimiento de la divulgación para reducir la unión alejada del objetivo y los eventos de escisión. Un ácido nucleico objetivo (Fig. 5, 215) se pone en contacto con un complejo que comprende un polipéptido dirigido al sitio (por ejemplo, una Cas9 activa) (Fig. 5, 205) y un ácido nucleico que se dirige al ácido nucleico (por ejemplo, ARNgu o ARN guía doble) (Fig. 5, 210). El complejo se une a un ácido nucleico objetivo (Fig. 5, 220). En algunos casos, el complejo que comprende Cas9 y ARNgu se une a un ácido nucleico alejado del objetivo (Fig. 5, 225). Los complejos que comprenden una proteína dCas9 modificada (Fig. 5, 235) y un ARNgu modificado (Fig. 5, 236), pueden desplazarse y ponerse en contacto (Fig. 5, 230) con el ácido nucleico objetivo. Los complejos de dCas9 pueden desplazar o prevenir la unión de complejos que comprenden Cas9 activa. La Cas9 activa puede escindir (Fig. 5, 240/245) el ácido nucleico objetivo. Se evita que la Cas9 activa escinda el ácido nucleico alejado del objetivo porque la dCas9 está impidiendo su unión y escisión. De esta manera, se puede prevenir la escisión no deseada.

La Fig. 6 muestra el uso de nickasas Cas9 D10A en tándem para cortar una región de cadena sencilla de ADN en la misma cadena de un ácido nucleico objetivo e insertar un polinucleótido donante. La Fig. 6A muestra dos complejos de ARNgu/dCas9 D10A dirigidos a dos posiciones separadas en la cadena sentido de un polinucleótido objetivo. La Fig. 6B muestra que una región en la cadena sentido entre los sitios objetivo ha sido escindida. La Fig. 6C muestra la inserción del polinucleótido donante con regiones flanqueantes superpuestas.

La Fig. 7A y la Fig. 7B representan procedimientos para aumentar la HDR utilizando ARNgu/dCas9 y complejos ARNgu /Cas9 catalíticamente activos. La Fig. 7A muestra un sistema que usa un solo complejo de ARNgu/dCas9 atado a un polinucleótido donante de HDR adyacente a un complejo de ARNgu/Cas9 activo para dirigir al donante al sitio del rompimiento de cadena doble y colocar al donante al lado del sitio de corte. La Fig. 7B muestra un sistema que utiliza dos complejos ARNgu/dCas9 separados y un complejo de ARNgu/Cas9 catalíticamente activo colocado entre los dos complejos inactivos catalíticamente, en el que el donante se coloca a través del rompimiento de cadena doble.

La Fig. 8 muestra los efectos de los bloqueadores de la nucleasa dCas9 (dCas9-NB) en la edición deseada de ARNgu/ Cas9 del *VEGFA* en el locus del *VEGFA*.

La Fig. 9 muestra los efectos de dCas9-NB en la edición no deseada de ARNgu/Cas9 del *VEGFA* en un locus alejado del objetivo conocido del *VEGFA* en el cromosoma 15 humano.

La Fig. 10 muestra los diversos aspectos de las condiciones experimentales utilizadas para colocar los nucleótidos del donante de homología cerca de un sitio objetivo para aumentar la eficiencia de HDR, como se describe en el Ejemplo 5C.

La Fig. 11 muestra las configuraciones de donantes potenciales utilizando Cas9D10A en tándem como se describe en los ejemplos.

La Fig. 12 muestra una comparación de tipos de reparación utilizando Cas9 o Cas9D10A en los Objetivos 3 y 4 (locus CD34 humano) como se describe en los ejemplos. Neg denota un control negativo que es solo Cas9 o Cas9D10A, sin ARNgu. La distribución de reparación se indica mediante las barras en la figura. Barras continuas = sin editar; barras rayadas = reparación mutagénica; barras punteadas = HDR.

Descripción detallada de la invención

La invención se define por las reivindicaciones adjuntas. Debe entenderse que la terminología utilizada en este documento tiene el propósito de describir solo realizaciones particulares, y no pretende ser limitante. Tal como se utiliza en esta memoria descriptiva y en las reivindicaciones adjuntas, las formas singulares "un", "uno, una" y "el, la" incluyen referencias al plural, a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "un complejo de ARNgu/dCas9" incluye uno o más de estos complejos, la referencia a "un complejo de ARNgu/Cas9" incluye uno o más de estos complejos, la referencia a "una mutación" incluye una o más mutaciones, y similares. También debe entenderse que cuando se hace referencia a una realización que usa un ARNgu para dirigir Cas9 o dCas9 a un sitio objetivo, un experto en la técnica puede usar una realización alternativa de la invención con base en el uso de un ARN guía dual (por ejemplo, ARNcr/ARNcrtra) en lugar del ARNgu.

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que entiende comúnmente un experto en la técnica a la que pertenece la invención. Aunque en la práctica de la presente invención se pueden usar otros procedimientos y materiales similares o equivalentes a los descritos en el presente documento, los materiales y procedimientos preferidos se describen en el presente documento.

En vista de las enseñanzas de la presente memoria descriptiva, un experto en la materia puede aplicar técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y polinucleótidos recombinantes, como se enseña, por ejemplo, mediante los siguientes textos estándar: *Antibodies: A Laboratory Manual*, Segunda edición, E. A. Greenfield, 2014, Cold Spring Harbor Laboratory Press, ISBN 978-1-936113-81-1; *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6a Edición, R. I. Freshney, 2010, Wiley-Blackwell, ISBN 978-0-470-52812-9; *Transgenic Animal Technology*, 3a Edición: A Laboratory

Handbook, 2014, C.A. Pinkert, Elsevier, ISBN 978-0124104907; The Laboratory Mouse, Segunda Edición, 2012, H. Hedrich, Academic Press, ISBN 978-0123820082; Manipulating the Mouse Embryo: A Laboratory Manual, 2013, R. Behringer, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1936113019; PCR 2: A Practical Approach, 1995, M. J. McPherson, et al., IRL Press, ISBN 978-0199634248; Methods in Molecular Biology (Series), J.M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, 2010, D. C. Rio, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879698911; Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (4a Edición), 2012, M.R. Green, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1605500560; Bioconjugate Techniques, 3a Edición, 2013, G. T. Hermanson, Academic Press, ISBN 978-0123822390; Methods in Plant Biochemistry and Molecular Biology, 1997, W.V. Dashek, CRC Press, ISBN 978-0849394805; Plant Cell Culture Protocols (Methods in Molecular Biology), 2012, V. M. Loyola-Vargas, et al., Humana Press, ISBN 978-1617798177; Plant Transformation Technologies, 2011, C. N. Stewart, et al., Wiley-Blackwell, ISBN 978-0813821955; Recombinant Proteins from Plants (Methods in Biotechnology), 2010, C. Cunningham, et al., Humana Press, ISBN 978-1617370212; Plant Genomics: Methods and Protocols (Methods in Molecular Biology), 2009, D. J. Somers, et al., Humana Press, ISBN 978-1588299970; Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, 2008, R. Keshavachandran, et al., Orient Blackswan, ISBN 978-8173716164.

El término "proteína Cas9" como se usa en este documento se refiere a proteínas CRISPR-Cas9 Tipo II (como se describe, por ejemplo, en Chylinski, K., (2013) "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems", RNA Biol. 2013 10 (5): 726-737), que incluye, entre otros, proteínas Cas9, similares a Cas9, codificadas por ortólogos de Cas9, proteínas sintéticas de tipo Cas9 y sus variantes y modificaciones. El término "proteína Cas9", como se usa en el presente documento, se refiere a las proteínas Cas9 de tipo silvestre derivadas de los sistemas CRISPR-Cas9 de Tipo II, a las modificaciones de las proteínas Cas9, a las variantes de las proteínas Cas9, a los ortólogos de Cas9 y a sus combinaciones. Las proteínas Cas9 pueden derivarse de cualquiera de las diversas especies bacterianas cuyos genomas codifican tales proteínas. Las proteínas Cas para uso en los presentes procedimientos se describen más adelante.

Los términos "de tipo silvestre", "de origen natural" y "sin modificar" se usan en el presente documento para referirse a la forma, apariencia, fenotipo o cepa típica (o más común) que existe en la naturaleza; por ejemplo, la forma típica de células, organismos, características, polinucleótidos, proteínas, complejos macromoleculares, genes, ARN, ADN o genomas a medida que aparecen y se pueden aislar de una fuente en la naturaleza. La forma, la apariencia, el fenotipo o la cepa de tipo silvestre sirven como el compuesto original antes de una modificación intencional. Por lo tanto, las formas mutantes, variantes, manipuladas, recombinantes y modificadas no son formas de tipo silvestre.

Como se usa en el presente documento, los términos "manipulado", "modificado genéticamente", "recombinante", "modificado" y "de origen no natural" son intercambiables e indican manipulación humana intencional.

Como se usa en el presente documento, los términos "ácido nucleico", "secuencia de nucleótidos", "oligonucleótido" y "polinucleótido" son intercambiables. Todos se refieren a una forma polimérica de nucleótidos. Los nucleótidos pueden ser desoxirribonucleótidos (ADN) o ribonucleótidos (ARN), o análogos de los mismos, y pueden ser de cualquier longitud. Los polinucleótidos pueden realizar cualquier función y pueden tener cualquier estructura secundaria y estructura tridimensional. Los términos abarcan análogos conocidos de nucleótidos naturales y nucleótidos que se modifican en las fracciones de la base, el azúcar y/o el fosfato. Los análogos de un nucleótido particular tienen la misma especificidad de apareamiento de bases (por ejemplo, un análogo de los pares de bases A con T). Un polinucleótido puede comprender un nucleótido modificado o múltiples nucleótidos modificados. Los ejemplos de nucleótidos modificados incluyen nucleótidos metilados y análogos de nucleótidos. La estructura del nucleótido puede modificarse antes o después de ensamblar un polímero. Después de la polimerización, los polinucleótidos pueden modificarse adicionalmente, por ejemplo, mediante conjugación con un componente de marcaje o componente de unión a objetivo. Una secuencia de nucleótidos puede incorporar componentes no nucleotídicos. Los términos también abarcan ácidos nucleicos que comprenden residuos o enlaces de la cadena principal modificada, que (i) son sintéticos, naturales y no naturales, y (ii) tienen propiedades de unión similares a las de un polinucleótido de referencia (por ejemplo, ADN o ARN). Los ejemplos de dichos análogos incluyen, pero no se limitan a, fosforotioatos, fosforamidatos, fosfonatos de metilo, fosfonatos de metilo quirales, ribonucleótidos 2-O-metilo, ácidos nucleicos-péptidos (PNA) y estructuras morfolino.

Las secuencias de polinucleótidos se muestran en el presente documento en la orientación convencional de 5' a 3'.

Como se usa en el presente documento, el término "complementariedad" se refiere a la capacidad de una secuencia de ácido nucleico para formar un enlace o enlaces de hidrógeno con otra secuencia de ácido nucleico (por ejemplo, a través del emparejamiento tradicional de bases de Watson-Crick). Un porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que puede formar enlaces de hidrógeno con una segunda secuencia de ácido nucleico. Cuando dos secuencias de polinucleótidos tienen una complementariedad del 100%, las dos secuencias son perfectamente complementarias, es decir, todos los enlaces de hidrógeno de residuos contiguos del primer polinucleótido con el mismo número de residuos contiguos en un segundo polinucleótido.

Como se usa en el presente documento, el término "identidad de secuencia" generalmente se refiere al porcentaje de identidad de bases o aminoácidos determinado comparando un primer polinucleótido o polipéptido con un segundo polinucleótido o polipéptido usando algoritmos que tienen varios parámetros de ponderación. La identidad

de la secuencia entre dos polipéptidos o dos polinucleótidos se puede determinar mediante el alineamiento de secuencias mediante diversos procedimientos y programas informáticos (por ejemplo, BLAST, CS-BLAST, FASTA, HMMER, L-ALIGN, etc.), disponibles a través de la red mundial en sitios que incluyen GENBANK (ncbi.nlm.nih.gov/genbank/) y EMBL-EBI (ebi.ac.uk.). La identidad de secuencia entre dos polinucleótidos o dos secuencias polipeptídicas generalmente se calcula utilizando los parámetros estándar predeterminados de los diversos procedimientos o programas informáticos.

Como se usa en este documento, una "estructura de tallo-bucle" o "elemento de tallo-bucle" se refiere a un polinucleótido que tiene una estructura secundaria que incluye una región de nucleótidos que se sabe o predice que forman una región de cadena doble (el "elemento del tallo") que está enlazado por un lado por una región de nucleótidos predominantemente de cadena sencilla (el "elemento del bucle"). El término elemento de "horquilla" también se usa en el presente documento para referirse a estructuras de tallo-bucle. Tales estructuras son bien conocidas en la técnica. El emparejamiento de bases puede ser exacto. Sin embargo, como se conoce en la técnica, un elemento del tallo no requiere un emparejamiento de bases exacto. Por lo tanto, el elemento del tallo puede incluir uno o más desajustes de bases o bases no emparejadas.

Como se usa en el presente documento, el término "recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos.

Como se usa en el presente documento, el término "reparación dirigida por homología" o "HDR" se refiere a la reparación del ADN que tiene lugar en las células, por ejemplo, durante la reparación de rompimientos de cadena doble y de cadena sencilla en el ADN. La HDR requiere una homología de secuencia de nucleótidos y utiliza una "plantilla de donante" (ADN plantilla del donante, polinucleótido donante u oligonucleótido (se usa de manera intercambiable en el presente documento) para reparar la secuencia donde ocurrió el rompimiento de cadena doble (por ejemplo, secuencia diana ADN). Este resulta en la transferencia de información genética de, por ejemplo, el ADN de la plantilla del donante a la secuencia de ADN diana. La HDR puede producir una alteración de la secuencia de ADN diana (por ejemplo, inserción, eliminación, mutación) si la secuencia del ADN de la plantilla del donante o la secuencia de oligonucleótidos difiere de la secuencia de ADN diana y parte o la totalidad del polinucleótido u oligonucleótido de ADN de la plantilla del donante se incorporan en la secuencia de ADN diana. En algunos aspectos, un polinucleótido de ADN de la plantilla del donante completo, una parte del polinucleótido de ADN de la plantilla del donante, o una copia del polinucleótido del donante está integrado en el sitio de la secuencia de ADN diana.

Como se usa en el presente documento, el término "unión de extremos no homólogos" o "NHEJ" se refiere a la reparación de rompimientos de cadena doble en el ADN mediante la ligadura directa de un extremo del rompimiento al otro extremo del rompimiento sin un requisito por un ADN plantilla del donante. La NHEJ en ausencia de un ADN de plantilla del donante a menudo da como resultado un pequeño número de nucleótidos insertados o eliminados al azar en el sitio del rompimiento de la cadena doble.

Los mecanismos alternativos de inserción de ADN que no requieren homología de secuencia entre el donante y la secuencia objetivo también pueden usarse para la inserción de ácido nucleico. Estos mecanismos involucran varios componentes de la maquinaria de reparación de ADN celular.

Los términos "vector" y "plásmido" se usan de manera intercambiable y, como se usa en el presente documento, se refieren a un vehículo polinucleótido para introducir material genético en una célula. Los vectores pueden ser lineales o circulares. Los vectores pueden integrarse en un genoma objetivo de una célula huésped o replicarse independientemente en una célula huésped. Los vectores pueden comprender, por ejemplo, un origen de replicación, un sitio de clonación múltiple y/o un marcador seleccionable. Un vector de expresión comprende típicamente un casete de expresión. Los vectores y plásmidos incluyen, pero no se limitan a, vectores de integración, plásmidos procariontes, plásmidos eucariotes, cromosomas sintéticos de plantas, episomas, vectores virales, cósmidos y cromosomas artificiales. Como se usa en el presente documento, el término "casete de expresión" es una construcción polinucleotídica, generada de forma recombinante o sintética, que comprende secuencias reguladoras unidas operativamente a un polinucleótido seleccionado para facilitar la expresión del polinucleótido seleccionado en una célula huésped. Por ejemplo, las secuencias reguladoras pueden facilitar la transcripción del polinucleótido seleccionado en una célula huésped, o la transcripción y traducción del polinucleótido seleccionado en una célula huésped. Un casete de expresión puede, por ejemplo, integrarse en el genoma de una célula huésped o estar presente en un vector de expresión.

Como se usa en el presente documento, el término "casete de expresión" es una construcción polinucleotídica, generada de forma recombinante o sintética, que comprende secuencias reguladoras unidas operativamente a un polinucleótido seleccionado para facilitar la expresión del polinucleótido seleccionado en una célula huésped. Por ejemplo, las secuencias reguladoras pueden facilitar la transcripción del polinucleótido seleccionado en una célula huésped, o la transcripción y traducción del polinucleótido seleccionado en una célula huésped. Un casete de expresión puede, por ejemplo, integrarse en el genoma de una célula huésped o estar presente en un vector de expresión.

Como se usa en el presente documento, un "vector de direccionamiento" es un constructo de ADN recombinante

que típicamente comprende brazos de ADN adaptados y homólogos al ADN genómico que flanquea los elementos críticos de un gen o secuencia objetivo. Cuando se introduce en una célula, el vector de direccionamiento se integra en el genoma de la célula a través de una recombinación homóloga. Los elementos del gen objetivo se pueden modificar de varias maneras, incluyendo eliminaciones y/o inserciones. Un gen objetivo defectuoso puede ser reemplazado por un gen objetivo funcional, o en forma alternativa, un gen funcional puede ser eliminado. Opcionalmente, un vector de direccionamiento comprende un casete de selección que comprende un marcador seleccionable que se introduce en el gen objetivo. Las regiones de direccionamiento adyacentes o, a veces, dentro de un gen objetivo se pueden usar para afectar la regulación de la expresión génica.

Como se usa en el presente documento, los términos "secuencias reguladoras", "elementos reguladores" y "elementos de control" son intercambiables y se refieren a secuencias de polinucleótidos que están en sentido 5' (secuencias no codificantes 5'), dentro o en sentido 3' (secuencias no traducidas 3') de un objetivo de polinucleótido a expresar. Las secuencias reguladoras influyen, por ejemplo, en la sincronización de la transcripción, la cantidad o el nivel de transcripción, el procesamiento o la estabilidad del ARN y/o la traducción de la secuencia relacionada de nucleótidos estructurales. Las secuencias reguladoras pueden incluir secuencias de unión a activador, potenciadores, intrones, secuencias de reconocimiento de poliadenilación, promotores, secuencias de unión a represor, estructuras de tallo-bucle, secuencias de iniciación de la traducción, secuencias líderes de la traducción, secuencias de terminación de la transcripción, secuencias de terminación de la traducción, sitios de unión del cebador, y similares.

Como se usa en el presente documento, el término "unido operativamente" se refiere a secuencias de polinucleótidos o secuencias de aminoácidos colocadas en una relación funcional entre sí. Por ejemplo, un promotor o potenciador está unido operativamente a una secuencia de codificación si regula, o contribuye a la modulación de la transcripción de la secuencia de codificación. Las secuencias de ADN unidas operativamente que codifican secuencias reguladoras son típicamente contiguas a la secuencia codificante. Sin embargo, los potenciadores pueden funcionar cuando están separados de un promotor por hasta varias kilobases o más. Por consiguiente, algunos elementos polinucleotídicos pueden estar operativamente unidos pero no son contiguos.

Como se usa en el presente documento, el término "expresión" se refiere a la transcripción de un polinucleótido de un molde de ADN, dando como resultado, por ejemplo, un ARNm u otro transcrito de ARN (por ejemplo, no codificante, tal como los ARN estructurales o de andamiaje). El término se refiere además al proceso a través del cual el ARNm transcrito se traduce en péptidos, polipéptidos o proteínas. Los transcritos y los polipéptidos codificados se pueden denominar colectivamente como "producto genético". La expresión puede incluir el empalme del ARNm en una célula eucariótica, si el polinucleótido se deriva del ADN genómico.

Como se usa en el presente documento, el término "modular" se refiere a un cambio en la cantidad, el grado o la cantidad de una función. Por ejemplo, los procedimientos descritos en el presente documento pueden modular la eficacia del direccionamiento mediado por Cas9 al disminuir o eliminar la escisión no deseada, mejorando así la división en el sitio objetivo, o pueden mejorar la HDR y disminuir la probabilidad de eventos NHEJ. Por consiguiente, el término "direccionamiento por modulación" puede indicar un aumento de los eventos de direccionamiento deseados y/o la inhibición de la escisión no deseada. De manera similar, "HDR de modulación" puede denotar un aumento de HDR y/o una disminución de NHEJ.

La modulación puede analizarse determinando cualquier característica directa o indirectamente afectada por la expresión del gen objetivo. Dichas características incluyen, por ejemplo, cambios en la eficacia de direccionamiento, niveles de ARN o de proteínas, actividad de las proteínas, niveles de productos, expresión de genes asociados o nivel de actividad de genes informadores. Por lo tanto, la "modulación" de la expresión génica incluye tanto la activación génica como la represión génica.

Como se usa en el presente documento, el término "aminoácido" se refiere a aminoácidos naturales y sintéticos (no naturales), que incluyen análogos de aminoácidos, aminoácidos modificados, peptidomiméticos, glicina e isómeros ópticos D o L.

Como se usa en el presente documento, los términos "péptido", "polipéptido" y "proteína" son intercambiables y se refieren a polímeros de aminoácidos. Un polipéptido puede ser de cualquier longitud. Puede ser ramificado o lineal, puede estar interrumpido por no aminoácidos y puede comprender aminoácidos modificados. Los términos se pueden usar para referirse a un polímero de aminoácidos que se ha modificado, por ejemplo, mediante acetilación, formación de enlaces disulfuro, glicosilación, lipidación, fosforilación, entrecruzamiento y/o conjugación (por ejemplo, con un componente de marcación o ligando). Las secuencias de polipéptidos se muestran en este documento en la orientación convencional del extremo terminal N al extremo terminal C.

Los polipéptidos y polinucleótidos se pueden elaborar usando técnicas de rutina en el campo de la biología molecular (véase, por ejemplo, los textos estándar discutidos anteriormente). Además, esencialmente cualquier polipéptido o polinucleótido puede ordenarse por costumbre a partir de fuentes comerciales.

El término "unión" como se usa en este documento incluye una interacción no covalente entre macromoléculas (por ejemplo, entre una proteína y un polinucleótido, entre un polinucleótido y un polinucleótido, y entre una proteína y

una proteína). Dicha interacción no covalente también se conoce como "asociación" o "interacción" (por ejemplo, cuando una primera macromolécula interactúa con una segunda macromolécula, la primera macromolécula se une a la segunda macromolécula de una manera no covalente). Algunas porciones de una interacción de unión pueden ser específicas de la secuencia; sin embargo, no es necesario que todos los componentes de una interacción de unión sean específicos de la secuencia, como los contactos de una proteína con residuos de fosfato en una cadena principal de ADN. Las interacciones de unión pueden caracterizarse por una constante de disociación (Kd). "Afinidad" se refiere a la fuerza de unión. Una mayor afinidad de unión se correlaciona con una Kd menor. Un ejemplo de unión no covalente es la formación de enlaces de hidrógeno entre pares de bases.

Como se usa en el presente documento, el término "aislado" puede referirse a un ácido nucleico o polipéptido que, por acción humana, existe aparte de su entorno nativo y, por lo tanto, no es un producto de la naturaleza. Aislado significa sustancialmente puro. Un ácido nucleico o polipéptido aislado puede existir en una forma purificada y/o puede existir en un entorno no nativo tal como, por ejemplo, en una célula recombinante.

Como se usa en este documento, una "célula huésped" generalmente se refiere a una célula biológica. Una célula puede ser la unidad básica estructural, funcional y/o biológica de un organismo vivo. Una célula puede originarse de cualquier organismo que tenga una o más células. Los ejemplos de células huésped incluyen, pero no se limitan a: una célula procariota, una célula eucariota, una célula bacteriana, una célula arqueal, una célula de un organismo eucariota de una sola célula, una célula protozoaria, una célula de una planta (por ejemplo, células de cultivos de plantas, frutas, vegetales, granos, soja, maíz, trigo, semillas, tomates, arroz, yuca, caña de azúcar, girasol, sorgo, mijo, alfalfa, *Brassica* productora de aceite (por ejemplo, pero no limitado a, colza oleaginosa / canola), calabaza, heno, patatas, algodón, cannabis, tabaco, plantas con flores, coníferas, gimnospermas, helechos, licopodios, hornabeques, hepáticas, musgos), una célula de alga, (por ejemplo, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* C. Agardh, y similares), algas marinas (por ejemplo, quelpos), una célula fúngica (por ejemplo, una célula de levadura, una célula de un hongo), una célula animal, una célula de un animal invertebrado (por ejemplo, mosca de la fruta, cnidaria, equinodermo, nematodo, etc.), una célula de un animal vertebrado (por ejemplo, pez, anfibio, reptil, ave, mamífero), una célula de un mamífero (por ejemplo, un cerdo, una vaca, una cabra, una oveja, un roedor, una rata, un ratón, un primate no humano, un humano, etc.). Además, una célula puede ser una célula madre o una célula progenitora.

Como se usa en el presente documento, el término "organismo transgénico" se refiere a un organismo que comprende un polinucleótido introducido de forma recombinante.

Como se usa en el presente documento, los términos "célula vegetal transgénica" y "planta transgénica" son intercambiables y se refieren a una célula vegetal o una planta que contiene un polinucleótido introducido de forma recombinante. En el término planta transgénica se incluye la progenie (cualquier generación) de una planta transgénica o una semilla, de manera que la progenie o semilla comprende una secuencia de ADN que codifica un polinucleótido introducido de forma recombinante o un fragmento de la misma.

Como se usa en el presente documento, la frase "generar una célula vegetal transgénica o una planta" se refiere a usar procedimientos y técnicas de ADN recombinante para construir un vector para la transformación de la planta para transformar la célula vegetal o la planta y para generar la célula vegetal transgénica o la planta transgénica.

CRISPR (repeticiones palindrómicas cortas agrupadas, regularmente espaciadas) es un locus genómico encontrado en los genomas de muchos procariotas (por ejemplo, bacterias y arqueas). Los loci de CRISPR proporcionan resistencia a invasores foráneos (por ejemplo, virus, fagos) en procariotas. De esta manera, se puede pensar que el sistema CRISPR funciona como un tipo de sistema inmunológico que ayuda a defender a los procariotas contra invasores foráneos. Hay tres etapas de la función del locus de CRISPR: integración de nuevas secuencias en el locus, biogénesis del ARN de CRISPR (ARNcr) y silenciamiento del ácido nucleico invasor foráneo.

Un locus de CRISPR incluye una serie de secuencias de repetición cortas denominadas "repeticiones". Las repeticiones pueden formar estructuras de horquilla y/o las repeticiones pueden ser secuencias de cadena sencilla no estructuradas. Las repeticiones ocurren en grupos. Las repeticiones con frecuencia divergen entre las especies. Las repeticiones se intercalan regularmente con secuencias intermedias únicas, denominadas "espaciadoras", lo que da como resultado una arquitectura de locus repetición-espaciadora-repetición. Las espaciadoras son idénticas o tienen una alta homología con secuencias de invasoras foráneas conocidas. Una unidad espaciadora-repetición codifica un ARNcr (ARNcr). Un ARNcr se refiere a la forma madura de la unidad de espaciadora-repetición. Un ARNcr comprende una secuencia "semilla" que participa en la selección de un ácido nucleico objetivo (por ejemplo, posiblemente como un mecanismo de vigilancia contra un ácido nucleico foráneo). Una secuencia semilla se ubica típicamente hacia el extremo 5' de un ARNcr (por ejemplo, en el complejo de Cascada; para una descripción del complejo de Cascada, véase, por ejemplo, Jore, MM et al., "Structural basis for CRISPR RNA-guided DNA recognition by Cascade", *Nature Structural & Molecular Biology* (2011) 18: 529-536) o en el extremo 3' del espaciador de un ARNcr (por ejemplo, en un sistema CRISPR-Cas9 Tipo II), directamente adyacente al primer tallo.

Un locus de CRISPR comprende secuencias de polinucleótidos que codifican genes de genes asociados a CRISPR (Cas). Los genes Cas están implicados en la biogénesis y/o las etapas de interferencia de la función de ARNcr. Los genes Cas muestran una extrema divergencia de secuencia (por ejemplo, una secuencia primaria) entre especies y

homólogos. Por ejemplo, los homólogos de Cas1 pueden comprender menos del 10% de identidad de secuencia primaria entre homólogos. Algunos genes Cas comprenden estructuras secundarias y/o terciarias homólogas. Por ejemplo, a pesar de la extrema divergencia de secuencias, muchos miembros de la familia Cas6 de proteínas CRISPR comprenden un pliegue similar a la ferredoxina del extremo terminal N. Los genes Cas se nombran de acuerdo con el organismo del cual se derivan. Por ejemplo, los genes Cas en *Staphylococcus epidermidis* pueden denominarse de tipo Csm, los genes Cas en *Streptococcus thermophilus* se pueden denominar de tipo Csn y los genes Cas en *Pyrococcus furiosus* se pueden denominar de tipo Cmr.

La etapa de integración de un sistema CRISPR se refiere a la capacidad del locus de CRISPR para integrar nuevos espaciadores en la matriz de ARNcr tras ser infectado por un invasor foráneo. La adquisición de espaciadores de invasores foráneos puede ayudar a conferir inmunidad a ataques subsiguientes por parte del mismo invasor foráneo. La integración ocurre típicamente en el extremo líder del locus de CRISPR. Las proteínas Cas (por ejemplo, Cas1 y Cas2) están involucradas en la integración de nuevas secuencias espaciadoras. La integración se realiza de manera similar para algunos tipos de sistemas CRISPR (por ejemplo, Tipo I-III).

Los ARNcr maduros se procesan a partir de un transcrito de locus de CRISPR policistrónico más largo (es decir, una matriz de pre-ARNcr). Una matriz pre-ARNcr comprende una pluralidad de ARNcr. Las repeticiones en la matriz de pre-ARNcr son reconocidas por los genes Cas. Los genes Cas se unen a las repeticiones y escinden las repeticiones. Esta acción puede liberar la pluralidad de los ARNcr. Los ARNcr pueden someterse a eventos adicionales para producir la forma de ARNcr madura, como un recorte (por ejemplo, con una exonucleasa). Un ARNcr puede comprender todas, algunas o ninguna de las secuencias de repetición CRISPR.

La interferencia se refiere a la etapa en el sistema de CRISPR que es funcionalmente responsable de combatir la infección por un invasor foráneo. La interferencia de CRISPR sigue un mecanismo similar al ARN de interferencia (ARNi: por ejemplo, en el que un ARN objetivo es seleccionado (por ejemplo, hibridado) por un ARN de interferencia corto (ARNic)), que resulta en la degradación y/o desestabilización del ARN objetivo. Los sistemas CRISPR realizan la interferencia de un ácido nucleico objetivo mediante el acoplamiento de los ARNcr y los genes Cas, formando así las ribonucleoproteínas CRISPR (RNPcr). El ARNcr de RNPcr guía al RNPcr hacia un ácido nucleico invasor foráneo (por ejemplo, reconociendo el ácido nucleico invasor foráneo a través de la hibridación). Las unidades de ácido nucleico invasor foráneo-ARNcr hibridadas se someten a escisión por proteínas Cas. La interferencia del ácido nucleico objetivo normalmente requiere un motivo adyacente protoespaciador (PAM) en un ácido nucleico objetivo.

Hay al menos cuatro tipos de sistemas CRISPR: Tipo I, Tipo II, Tipo III y Tipo U. Se puede encontrar más de un sistema de tipo CRISPR en un organismo. Los sistemas CRISPR pueden ser complementarios entre sí, y/o pueden prestar unidades funcionales en trans para facilitar el procesamiento del locus de CRISPR. Los sistemas Tipo II se pueden subdividir además en II-A (contiene el locus Csn2) y II-B (contiene el locus Cas4) y el Tipo II-C (ni Csn2 ni Cas4, por ejemplo, *N. meningitidis*). Las modificaciones de los componentes de los sistemas CRISPR Tipo II se discuten ampliamente en la presente memoria descriptiva.

La biogénesis de ARNcr en un sistema CRISPR de Tipo II comprende un ARN de CRISPR de activación trans (ARNcrtra). Un ARNcrtra se modifica típicamente por RNAsaIII endógena. El ARNcrtra hibrida con una repetición de ARNcr en la matriz pre-ARNcr. Se recluta la RNAsaIII endógena para escindir el pre-ARNcr. Los ARNcr escindidos se someten a un recorte por exorribonucleasa para producir la forma de ARNcr madura (por ejemplo, recorte en 5'). El ARNcrtra típicamente permanece hibridado con el ARNcr. El ARNcrtra y el ARNcr se asocian con un polipéptido dirigido al sitio (por ejemplo, Cas9). El ARNcr del complejo de ARNcr-ARNcrtra-Cas9 puede guiar el complejo a un ácido nucleico objetivo con el que el ARNcr puede hibridar. La hibridación del ARNcr con el ácido nucleico objetivo activa una Cas9 afín de tipo silvestre para la escisión del ácido nucleico objetivo. El ácido nucleico objetivo en un sistema CRISPR de Tipo II comprende un PAM. En algunas realizaciones, un PAM es esencial para facilitar la unión de un polipéptido dirigido al sitio (por ejemplo, Cas9) a un ácido nucleico objetivo.

Cas9 es un ejemplo de proteína Cas de CRISPR de tipo II. Cas9 es una endonucleasa que puede ser programada por el ARNcrtra/ARNcr para escindir, específicamente en el sitio, el ADN diana utilizando dos dominios de endonucleasa distintos (dominios similares a HNH y RuvC/RNasa H) (véase la solicitud de patente publicada de Estados Unidos No. 2014-0068797, publicada el 6 de marzo de 2014; véase también Jinek M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* (2012) 337: 816-821), una para cada cadena de la hélice doble de ADN. RuvC y HNH juntos producen rompimientos de cadena doble (DSB), y por separado pueden producir rompimientos de cadena sencilla. La Fig. 3A presenta un modelo de la disposición de dominio de SpyCas9 (Cas9 de *S. pyogenes*) en relación con su estructura de secuencia primaria. Dos componentes de ARN de un sistema CRISPR-Cas9 Tipo II se ilustran en la Fig. 1A. Típicamente, cada sistema CRISPR-Cas9 comprende un ARNcrtra y un ARNcr. Sin embargo, este requisito se puede omitir mediante el uso de un ARNcr modificado, que se describe más detalladamente a continuación, que contiene una horquilla diseñada que imita al complejo de ARNcrtra-ARNcr (Jinek et al., 2012). El emparejamiento de bases entre el ARNcr y el ADN diana causa rompimientos de cadena doble (DSB) debido a la actividad endonucleasa de Cas9. La especificidad de unión se determina mediante el emparejamiento de bases ARNcr-ADN y un motivo de ADN corto (secuencia del motivo adyacente protoespaciador [PAM]: NGG) yuxtapuesto a la región complementaria de DNA (Marraffini LA, Sontheimer EJ. "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea", *Nat Rev Genet.*, 2010; 11: 181-190). Por lo tanto, el sistema CRISPR solo requiere un conjunto mínimo de dos moléculas, la proteína

Cas9 y el ARN_g.

Un gran número de ortólogos de Cas9 son conocidos en la técnica, así como sus componentes ARN_{cr} y ARN_{tr} asociados (véase, por ejemplo, "Supplementary Table S2. List of bacterial strains with identified Cas9 orthologs", Fonfara, Ines, et al., "Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR/Cas Systems", *Nucleic Acids Research* (2014) 42: 2577-2590, incluidos todos los datos complementarios; Chylinski K., et al., "Classification and evolution of type II CRISPR-Cas systems", *Nucleic Acids Research* (2014) 42: 6091-6105, incluidos todos los datos complementarios.); Esvelt, K. M., et al., "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing", *Nature Methods* (2013) 10: 1116-1121). Se han identificado varias proteínas Cas9 ortogonales, incluidas las proteínas Cas9 de *Neisseria meningitidis*, *Streptococcus thermophilus* y *Staphylococcus aureus*.

Como se usa en el presente documento, "una proteína Cas9" se refiere a una proteína Cas9 derivada de cualquier especie, subespecie o cepa de bacterias que codifica Cas9, así como a las variantes y ortólogos de la Cas9 particular en cuestión. Las proteínas Cas9 se pueden aislar y purificar directamente a partir de bacterias, o producirse de forma sintética o recombinante, o se pueden administrar típicamente utilizando un constructo que codifica la proteína, incluyendo sin limitación, ADN desnudo, ADN plasmídico, un vector viral y ARN_m para la expresión de Cas9.

Las variantes y modificaciones de las proteínas Cas9 son conocidas en la técnica. La solicitud de patente publicada de Estados Unidos No. 20140273226, publicada el 18 de septiembre de 2014, analiza el gen Cas9 de *S. pyogenes*, la proteína Cas9 y las variantes de la proteína Cas9, incluidas las secuencias codificantes de Cas9 optimizadas en codones específicos del huésped (por ejemplo, 0129-0137 en este documento) y las proteínas de fusión Cas9 (por ejemplo, 233-240 en este documento). La solicitud de patente publicada de los Estados Unidos No. 20140315985, publicada el 23 de octubre de 2014, enseña un gran número de ejemplos de polipéptidos Cas9 de tipo silvestre (por ejemplo, las SEQ ID NOS: 1-256, SEQ ID NOS: 795-1346 en este documento), incluida la secuencia de Cas9 de *S. pyogenes* (SEQ ID NO: 8, en este documento). También se discuten las modificaciones y variantes de las proteínas Cas9 (por ejemplo, 504-608, en este documento). Ejemplos no limitantes de proteínas Cas9 incluyen proteínas Cas9 de *S. pyogenes* (GI: 15675041); *Listeria innocua* Clip 11262 (GI: 16801805); *Streptococcus mutans* UA159 (GI: 24379809); *Streptococcus thermophilus* LMD-9 (*S. thermophilus* A, GI:11662823; *S. thermophilus* B, GI:116627542); *Lactobacillus buchneri* NRRL B-30929 (GI:331702228); *Treponema denticola* ATCC 35405 (GI: 42525843); *Francisella novicida* U112 (GI: 118497352); *Campylobacter jejuni* subespecie *Jejuni* NCTC 11168 (GI: 218563121); *Pasteurella multocida* subespecie *multocida* str. Pm70 (GI:218767588); *Neisseria meningitidis* Zs491 (GI:15602992) y *Actinomyces naeslundii* (GI:489880078).

Un experto en la técnica puede poner en práctica los aspectos de la presente invención siguiendo la guía de la memoria descriptiva para usar proteínas CRISPR-Cas9 de Tipo II y polinucleótidos que codifican la proteína Cas, incluidos, entre otros, Cas9, similar a Cas9, las proteínas codificadas por los ortólogos de Cas9, las proteínas sintéticas similares a Cas9, y sus variantes y modificaciones. Los componentes de ARN afines de estas proteínas Cas pueden manipularse y modificarse para su uso en la práctica de la presente invención por un experto en la técnica siguiendo la guía de la presente memoria descriptiva.

Por "dCas9" se entiende una Cas9 desactivada por nucleasa, también denominada "catalíticamente inactiva", "Cas9 catalíticamente muerta" o "Cas9 muerta". Dichas moléculas carecen de toda o una parte de la actividad de la endonucleasa y, por lo tanto, pueden usarse para regular los genes de una manera guiada por el ARN (Jinek M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* (2012) 337: 816-821). Esto se logra mediante la introducción de mutaciones que inactivan la función de la nucleasa Cas9 y se logra típicamente mediante la mutación de los dos residuos catalíticos (D10A en el dominio RuvC-1, y H840A en el dominio HNH, numerados con respecto a Cas9 de *S. pyogenes*) del gen que codifica Cas9. Se entiende que la mutación de otros residuos catalíticos para reducir la actividad de cualquiera o ambos de los dominios de nucleasa también puede llevarse a cabo por un experto en la técnica. Al hacerlo, dCas9 no puede escindir el ADN_{cd} pero conserva la capacidad de dirigirse al ADN. El doble mutante de Cas9 con cambios en las posiciones de aminoácidos D10A y H840A desactiva completamente las actividades tanto de nucleasa como de nickasa. La especificidad del direccionamiento se determina mediante el emparejamiento de bases complementarias de un ARN_g para el locus genómico y el motivo adyacente protoespaciador (PAM).

dCas9 se puede usar solo o en fusiones para reprimir sintéticamente (CRISPRi) o activar (CRISPRa) la expresión del gen. CRISPRi puede trabajar independientemente de las maquinarias celulares del huésped. En algunas realizaciones, solo una proteína dCas9 y un ARN_g personalizado diseñado con una región complementaria a cualquier gen de interés dirigen dCas9 a una ubicación genómica elegida. En otros aspectos, dCas9 puede fusionarse con un factor de transcripción, tal como un represor, y el factor de transcripción de Cas9 fusionado puede funcionar junto con mecanismos celulares. La especificidad de unión se determina conjuntamente por la región complementaria en el ARN_g y un motivo de ADN corto (motivo adyacente protoespaciador o PAM) yuxtapuesto a la región complementaria del ADN, dependiente de la especie en cuestión (véase, por ejemplo, Anders C., et al. ., "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease", *Nature* (2014) 513: 569-573). En el caso de *S. pyogenes*, esta secuencia es NGG. Para lograr la represión transcripcional, dCas9 se puede usar por sí mismo (por lo que reprime la transcripción a través del impedimento estérico). Tomados en conjunto,

ARNgu y dCas9 proporcionan un sistema mínimo para la regulación de genes específicos en cualquier organismo. (Qi, L. S., et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression", *Cell* (2013) 152: 1173-1183). CRISPRa se realiza mediante fusiones del factor de transcripción (activador) de dCas9.

5 Por una "nickasa Cas9" se entiende un mutante de Cas9 que no conserva la capacidad de realizar rompimientos de cadena doble en una secuencia de ácido nucleico objetivo, pero mantiene la capacidad de unirse y realizar un rompimiento de cadena sencilla en un sitio objetivo. Tal mutante incluirá típicamente una mutación en uno, pero no ambos de los dominios de endonucleasa Cas9 (HNH y RuvC). Por lo tanto, una mutación de aminoácido en la posición D10A o H840A en Cas9, numerada con respecto a *S. pyogenes*, puede resultar en la inactivación de la actividad catalítica de la nucleasa y convertir Cas9 en una enzima nickasa que realiza rompimientos de una sola cadena en el sitio objetivo. Debe entenderse que otros polipéptidos dirigidos al sitio tales como meganucleasas, nucleasas TALE, nucleasas de dedos de cinc, MEGA-TAL y otros conocidos por los expertos en la técnica pueden usarse en aspectos alternativos.

15 ARNcr tiene una región de complementariedad a una secuencia potencial diana ADN (Fig. 1A, la región 5' oscura del ARNcr) y una segunda región que forma enlaces de hidrógeno de pares de bases con el ARNcrtra para formar una estructura secundaria, típicamente para formar al menos una estructura de tallo (Fig. 1A, la región clara que se extiende hasta el extremo 3' del ARNcr). La región de complementariedad a la diana ADN es el espaciador. El ARNcrtra y el ARNcr interactúan a través de varios enlaces de hidrógeno de pares de bases para formar estructuras de ARN secundarias, por ejemplo, como se ilustra en la Fig. 1B. La formación de complejos entre el ARNcrtra / ARNcr y la proteína Cas da como resultado un cambio conformacional de la proteína Cas que facilita la unión al ADN, las actividades de endonucleasa de la proteína Cas y la escisión por la endonucleasa del ADN específica del sitio guiada por ARNcr. Para que un complejo de proteína Cas / ARNcrtra / ARNcr escinda una secuencia diana ADN, la secuencia diana ADN es adyacente a un motivo adyacente protoespaciador (PAM) afín.

25 El término "ARNgu" típicamente se refiere a un ARN guía único (es decir, una secuencia de polinucleótidos contigua única) que esencialmente comprende un ARNcr conectado en su extremo 3' al extremo 5' de un ARNcrtra a través de una secuencia "bucle" (véase, por ejemplo, la solicitud de patente publicada de los Estados Unidos No. 20140068797, publicada el 6 de marzo de 2014). ARNgu interactúa con una proteína Cas afín esencialmente como se describe para los polinucleótidos ARNcrtra / ARNcr, como se discutió anteriormente. Similar al ARNcr, el ARNgu tiene un espaciador, una región de complementariedad a una secuencia potencial diana ADN (Fig. 2, 201), adyacente a una segunda región que forma enlaces de hidrógeno de pares de bases que forman una estructura secundaria, típicamente una estructura de tallo (Fig. 2, 202, 203, 204, 205). El término incluye ARN guía único truncados (ARNgutru) de aproximadamente 17-18 nt. (Véase, por ejemplo, Fu, Y. et al., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs", *Nat Biotechnol.* (2014) 32: 279-284). El término también abarca ARNgu miniatura funcionales con características fungibles eliminadas, pero que retienen un módulo esencial y conservado denominado "de vínculo" ubicado en la porción de ARNgu que corresponde a ARNcrtra (no a ARNcr). Véase, por ejemplo, la solicitud de patente publicada de Estados Unidos No. 20140315985, publicada el 23 de octubre de 2014; Briner et al., "Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality", *Molecular Cell* (2014) 56: 333-339. El vínculo se encuentra inmediatamente secuencia abajo de (es decir, ubicado en la dirección 3' del) el tallo inferior en los sistemas CRISPR-Cas9 Tipo II. Un ejemplo de la ubicación relativa del vínculo se ilustra en el ARNgu que se muestra en la Fig. 2. El vínculo confiere la unión de un ARNgu o un ARNcrtra a su proteína Cas9 afín y confiere una apoenzima a la transición conformacional a haloenzima.

45 Con referencia a un ARNcr o ARNgu, un "espaciador" o "elemento espaciador" como se usa en el presente documento se refiere a la secuencia de polinucleótidos que puede hibridar específicamente con una secuencia de ácido nucleico objetivo. El elemento espaciador interactúa con la secuencia de ácido nucleico objetivo a través de enlaces de hidrógeno entre pares de bases complementarias (es decir, bases emparejadas). Un elemento espaciador se une a una secuencia diana ADN seleccionada. Por consiguiente, el elemento espaciador es una secuencia de unión a diana ADN. El elemento espaciador determina la ubicación de la unión específica de sitio de la proteína Cas y la escisión endonucleolítica. Los elementos espaciadores varían de ~17- a ~84 nucleótidos de longitud, dependiendo de la proteína Cas con la que están asociados, y tienen una longitud promedio de 36 nucleótidos (Marraffini, et al., "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea", *Nature reviews Genetics* (2010) 11: 181-190). En un sistema CRISPR-Cas9 Tipo II, el elemento espaciador típicamente comprende una secuencia "semilla" que está involucrada en la selección de un ácido nucleico objetivo. Por ejemplo, para SpyCas9, la longitud funcional de un espaciador para una escisión específica directa es típicamente de aproximadamente 12-25 nucleótidos. La variabilidad de la longitud funcional para un elemento espaciador es conocida en la técnica (por ejemplo, la solicitud de patente publicada de Estados Unidos No. 20140315985, publicada el 23 de octubre de 2014).

60 La Fig. 3A proporciona un modelo tridimensional basado en la estructura cristalina de Cas9 de *S. pyogenes* (SpyCas9) en un complejo activo con ARNgu. Se ilustra la relación del ARNgu con el dominio helicoidal y el dominio catalítico. Se indican los extremos 3' y 5' del ARNgu, así como las porciones expuestas del ARNgu. El ARN espaciador del ARNgu no es visible porque está rodeado por el lóbulo α -helicoidal (dominio helicoidal) y el lóbulo de nucleasa catalítica (dominio catalítico). El ARN espaciador del ARNgu se encuentra en la región del extremo 5' del ARNgu. Los dominios de nucleasa RuvC y HNH, cuando están activos, cortan una cadena de ADN diferente en el

ADN diana. El dominio del extremo terminal C (CTD) está involucrado en el reconocimiento de motivos adyacentes protoespaciadores (PAM) en el ADN diana.

La solicitud de patente publicada de Estados Unidos No. 20140315985, publicada el 23 de octubre de 2014; y Briner et al., "Guide RNA Modunc Functions Direct Cas9 Activity and Orthogonality", *Molecular Cell* (2014) 56: 333-339, divulgan secuencias de consenso y estructuras secundarias de ARN_g predichas para varias familias de ARN_g/Cas9. La disposición general de las estructuras secundarias en los ARN_g predichos arriba y que incluyen el vínculo se presentan en la Fig. 2 en este documento, que presenta una descripción general y una nomenclatura de los elementos del ARN_g de la Cas9 de *S. pyogenes*. En relación con la Fig. 2, existe una variación en el número y la disposición de las estructuras del tallo ubicadas 3' del vínculo en los ARN_g de la solicitud de patente publicada de Estados Unidos No. 2014-0315985 y Briner, et al.

Ran et al., "In vivo genome editing using *Staphylococcus aureus* Cas9", *Nature* (2015) 520: 186-191, que incluye todos los datos extendidos) presentan las secuencias ARN_{cr} / ARN_{cr}tr y estructuras secundarias de ocho sistemas CRISPR-Cas9 tipos II (véase la Figura 1 de Datos Extendidos de Ran, et al.). Además, Fonfara, et al., ("Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR/Cas Systems", *Nucleic Acids Research* (2014) 42: 2577-2590, incluidos todos los datos complementarios, en particular, la Figura Suplementaria S11) presenta las secuencias de ARN_{cr}/ARN_{cr}tr y las estructuras secundarias de ocho sistemas CRISPR-Cas9 Tipo II.

Por "polinucleótido de guía" se entiende cualquier polinucleótido que guíe específicamente al sitio a Cas9 o dCas9 a un ácido nucleico objetivo o alejado del objetivo. Se conocen muchos de estos polinucleótidos guía, tales como, pero sin limitación, ARN_g (incluyendo ARN_g miniatura y truncado), ARN guía dual, que incluye, entre otros, moléculas de ARN_{cr}/ARN_{cr}tr, como se describió anteriormente, y similares.

Por " polinucleótido donante" se entiende un polinucleótido que puede dirigirse e insertarse en un sitio objetivo de interés para modificar el ácido nucleico objetivo. Todo o una porción del polinucleótido donante puede insertarse en el ácido nucleico objetivo. El polinucleótido donante se usa para reparar el rompimiento en la secuencia de ADN diana que resulta en la transferencia de información genética (es decir, secuencias de polinucleótidos) del donante en el sitio o en la proximidad del rompimiento en el ADN. Por consiguiente, la nueva información genética (es decir, las secuencias de polinucleótidos) se puede insertar o copiar en un sitio de ADN diana. El polinucleótido donante puede ser ADN, ARN, un vector, un plásmido o similares, de cadena doble o de cadena sencilla. También se pueden usar donantes de polinucleótidos no simétricos que están compuestos por dos oligonucleótidos de ADN. Son parcialmente complementarios, y cada uno puede incluir una región flanqueante de homología. El donante se puede usar para insertar o reemplazar secuencias polinucleotídicas en una secuencia objetivo, por ejemplo, para introducir un polinucleótido que codifica una proteína o ARN funcional (por ejemplo, ARN_{ip}), para introducir una etiqueta de proteína, para modificar una secuencia reguladora a un gen, o para introducir una secuencia reguladora en un gen (por ejemplo, un promotor, un potenciador, una secuencia de entrada al ribosoma interno, un codón de inicio, un codón de parada, una señal de localización o una señal de poliadenilación), para modificar una secuencia de ácido nucleico (por ejemplo, introducir una mutación), y similares.

Las modificaciones de ADN dirigidas que usan polinucleótidos donantes para grandes cambios (por ejemplo, más de 100 pb de inserciones o supresiones) utilizan tradicionalmente plantillas de donantes basadas en plásmidos que contienen brazos de homología que flanquean el sitio de la alteración. Cada brazo puede variar en longitud, pero generalmente es más largo que alrededor de 100 pb, tal como 100-1.500 pb, por ejemplo, 100 ... 200 ... 300 ... 400 ... 500 ... 600 ... 700 ... 800 ... 900 ... 1.000 ... 1.500 pb o cualquier número entero entre estos valores. Sin embargo, estos números pueden variar, dependiendo del tamaño del polinucleótido donante y el polinucleótido objetivo. Este procedimiento se puede utilizar para generar grandes modificaciones, incluida la inserción de genes informadores, tales como las proteínas fluorescentes o los marcadores de resistencia a los antibióticos. Para la transfección en células, tales como las células HEK, aproximadamente 100-1.000 ng, por ejemplo, 100 ... 200 ... 300 ... 400 ... 500 ... 600 ... 700 ... 800 ... 900 ... 1.000 ng o cualquier número entero entre estos valores, de un plásmido donante de tamaño típico (por ejemplo, aproximadamente 5 kb) que contiene un vector ARN_g/Cas9, se puede usar para un pocillo en una placa de 24 pocillos. (Véase, por ejemplo, Yang et al., "One Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering" *Cell* (2013) 154: 1370-1379).

Se han usado oligonucleótidos de cadena sencillas y parcialmente de cadena doble, tales como oligonucleótidos de ADN, en lugar de plásmidos de direccionamiento para modificaciones cortas (por ejemplo, menos de 50 pb) dentro de un lugar definido sin clonación. Para lograr eficiencias altas de HDR, se pueden usar oligonucleótidos de cadena sencilla que contienen secuencias flanqueantes en cada lado que son homólogas a la región objetivo, y pueden orientarse en dirección sentido o antisentido en relación con el locus objetivo. La longitud de cada brazo puede variar en longitud, pero la longitud de al menos un brazo suele ser más larga que aproximadamente 10 bases, tal como de 10 a 150 bases, por ejemplo, 10 ... 20 ... 30 ... 40 ... 50 ... 60 ... 70 ... 80 ... 90 ... 100 ... 110 ... 120 ... 130 ... 140 ... 150, o cualquier número entero dentro de estos intervalos . Sin embargo, estos números pueden variar, dependiendo del tamaño del polinucleótido donante y el polinucleótido objetivo. En un aspecto preferido, la longitud de al menos un brazo es de 10 bases o más. En otros aspectos, la longitud de al menos un brazo es de 20 bases o más. En otros aspectos, la longitud de al menos un brazo es de 30 bases o más. En algunos aspectos, la longitud de al menos un brazo es inferior a 100 bases. En otros aspectos, la longitud de al menos un brazo es mayor que 100

bases. En algunos aspectos, la longitud de al menos un brazo es de cero bases. Para el diseño de oligonucleótidos de ADN de cadena sencilla, se usa típicamente un oligonucleótido con una homología total de alrededor de 100-150 pb. La mutación se introduce en el medio, produciendo brazos de homología de 50-75 pb para un donante diseñado para ser simétrico con respecto al sitio objetivo. En otros casos, no se requieren brazos de homología, y el polinucleótido donante se inserta utilizando mecanismos de reparación de ADN no homólogos.

En una realización, los procedimientos descritos en el presente documento son útiles para aumentar la eficiencia de la modificación mediada por Cas9 mediante la modulación de eventos de edición del genoma alejados del objetivo, por ejemplo, al disminuir el número de rompimientos de cadena doble en el ADN en ubicaciones no intencionadas y/o incorrectas. En particular, los sistemas de modificación del genoma, como los que usan nucleasas de dedos de zinc (ZFN), nucleasas TALE y nucleasas guiadas por ARN derivadas de bacterias (por ejemplo, el sistema CRISPR-Cas9), se han utilizado para dirigir una proteína a un locus genómico específico en el que puede inducir un rompimiento de cadena doble de ADN. Los rompimientos de cadena doble del ADN se pueden reparar a través de una unión de extremo no homólogo (NHEJ) o una reparación dirigida por homología (HDR). La NHEJ puede dar como resultado una reparación imperfecta y la adición o eliminación de varias bases, mientras que la HDR se puede utilizar para insertar secuencias de ADN exógenas modificadas racionalmente. Estos procedimientos a veces pueden resultar en una actividad de nucleasa no deseada como se describió anteriormente.

Los procedimientos para aumentar la especificidad y/o reducir los eventos genómicos alejados del objetivo han incluido el uso de secuencias guía más cortas con especificidad mejorada (Fu, Y. et al., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs", *Nat Biotechnol.* (2014) 32: 279-284) y/o modificación de mutantes de Cas9 que pueden usar dos eventos de direccionamiento independientes para inducir un rompimiento de cadena doble (Ran, FA, et al., "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity", *Cell* (2013) 154: 1380-1389; Tsai, SQ, et al., "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", *Nature Biotech.* (2014) 32: 569- 576). Sin embargo, estas estrategias pueden reducir la eficiencia de la edición del genoma objetivo, restringir las capacidades de direccionamiento o aún dar como resultado una actividad de nucleasa "alejada del objetivo".

Por consiguiente, una realización de la presente invención proporciona procedimientos para mitigar los eventos de edición del genoma alejado del objetivo en una población celular o en una reacción bioquímica *in vitro*. La mitigación de tales eventos puede ser realizada por un sistema CRISPR-Cas9 modificado como se describe en este documento. Los procedimientos incluyen al menos dos componentes básicos: (1) un complejo de una proteína Cas9 catalíticamente activa y un ARN_g que se dirige al objetivo de ácido nucleico objetivo (complejo de ARN_g/Cas9); y (2) un complejo de una proteína Cas9 catalíticamente inactiva, denominada "dCas9" en este documento y un ARN_g que se dirige a los loci alejados del objetivo (complejo de ARN_g/dCas9). En algunos aspectos, en lugar de un complejo de ARN_g/Cas9, el primer componente puede ser cualquier ADN endonucleasa de ADN catalíticamente activa dirigida al sitio, tal como, entre otros, nucleasas de dedos de zinc (ZFN), nucleasas TALE y similares.

Un ácido nucleico alejado del objetivo puede diferir de un ácido nucleico objetivo, por ejemplo, al menos 1-5, tal como 1, 2, 3, 4, 5 nucleótidos, o hasta 10 o más nucleótidos o cualquier número de nucleótidos dentro de los intervalos establecidos.

El porcentaje de complementariedad entre un locus de ácido nucleico alejado del objetivo (o región genómica circundante) y un ácido nucleico dirigido al ácido nucleico "objetivo" puede ser, por ejemplo, aproximadamente del 5% a aproximadamente el 100%, o cualquier porcentaje entre este intervalo, más preferiblemente en el intervalo de 90-100%.

Se conocen en la técnica varias proteínas Cas9 catalíticamente activas y, como se explicó anteriormente, una proteína Cas9 para su uso en el presente documento puede derivarse de cualquier especie bacteriana, subespecie o cepa que codifique la misma. Aunque la invención objeto se ejemplifica utilizando Cas9 de *S. pyogenes*, ortólogos de otras especies bacterianas encontrarán su uso en el presente documento. La especificidad de estos ortólogos de Cas9 es bien conocida. También son útiles las proteínas codificadas por proteínas sintéticas de tipo Cas9, y sus variantes y modificaciones. Como se explicó anteriormente, las secuencias para cientos de proteínas Cas9 son conocidas y cualquiera de estas proteínas encontrará uso con los procedimientos actuales. El experto en la técnica puede determinar fácilmente la proteína Cas9 adecuada para usar con un ácido nucleico objetivo particular.

Las proteínas dCas9 también son conocidas y, como se describió anteriormente, estas proteínas pueden volverse catalíticamente inactivas mediante la mutación de los dominios RuvC1 y/o HNH para eliminar la función nucleasa. Esto se logra típicamente mediante la introducción de mutaciones puntuales en los dos residuos catalíticos (D10A y H840A, numerados en relación con Cas9 de *S. pyogenes*) del gen que codifica Cas9. Al hacerlo, dCas9 se vuelve incapaz de escindir el ADN de cadena doble, pero conserva la capacidad de dirigirse al ADN. Además, al igual que con las proteínas Cas9, las proteínas dCas9 pueden derivarse de cualquier especie, subespecie o cepa bacteriana que codifique la misma. También son útiles las proteínas codificadas por los ortólogos de Cas9, las proteínas sintéticas de tipo Cas9 y sus variantes y modificaciones. En una realización, los ortólogos de dCas9 se seleccionan basándose en las secuencias particulares de motivos adyacentes protoespaciadores (PAM) presentes en el ácido nucleico objetivo. Por ejemplo, Cas9 de *S. pyogenes* apunta a secuencias NGG. Sin embargo, si otras secuencias de PAM están presentes, los ortólogos de dCas9 se pueden usar para dirigir estas secuencias para bloquear la

escisión de Cas9 de las mismas y evitar rompimientos alejados del objetivo.

En las siguientes realizaciones, el ARNgu se usa como un ejemplo de polinucleótido guía, sin embargo, un experto en la técnica reconocerá que se pueden usar otros polinucleótidos guía que guían específicamente al sitio a Cas9 o dCas9 a un ácido nucleico objetivo o alejado del objetivo. El componente de ARNgu de los complejos es responsable de apuntar a un objetivo de ácido nucleico particular. En particular, la región espaciadora del ARNgu incluye la región de complementariedad con la secuencia de ácido nucleico seleccionada. Por lo tanto, el espaciador es la secuencia polinucleotídica que puede hibridar específicamente con una secuencia de ácido nucleico objetivo. El elemento espaciador interactúa con la secuencia de ácido nucleico objetivo a través de enlaces de hidrógeno entre pares de bases complementarias. Un elemento espaciador se une a una secuencia objetivo de ácido nucleico seleccionada. Por consiguiente, el elemento espaciador es la secuencia de unión al diana ADN.

Por lo tanto, la especificidad de unión se determina tanto por el emparejamiento de bases ARNgu-ADN como por la secuencia de PAM yuxtapuesta a la región complementaria de ADN.

Por lo tanto, en un aspecto de la presente invención, un complejo de ARNgu/dCas9 se dirige a loci genómicos de manera similar a los complejos de ARNgu/Cas9 catalíticamente intactos, y puede unirse de manera estable al ADN y posteriormente bloquear la actividad de las proteínas dirigidas a esos loci. De esta manera, dCas9 puede alterar fuertemente la unión y/o la actividad de los factores de transcripción endógenos en células eucariotas.

En una realización a modo de ejemplo, un ARNgu, complejoado con Cas9 (complejo de ARNgu/Cas9) se dirige a un locus genómico de interés para inducir rompimientos de cadena doble. La especificidad de unión se determina mediante el emparejamiento de bases de ARNgu-DNA y la secuencia de PAM yuxtapuesta a la región complementaria de ADN. Se pueden usar procedimientos informáticos y/o experimentales (por ejemplo, secuenciación, procedimientos de alineación de ADN *in silico* para determinar la actividad de la nucleasa no deseada (por ejemplo, para determinar los loci alejados del objetivo). Estos procedimientos se describen detalladamente a continuación. Las proteínas dCas9 que actúan de manera independiente pueden diseñarse para dirigirse a estos loci alejados del objetivo. Estas proteínas dCas9 modificadas pueden implementarse como "bloqueadores" de la nucleasa específicos del sitio para obstruir catalíticamente la unión de ARNgu/Cas9 intacta y la actividad de nucleasa.

Se pueden introducir bloqueadores de ARNgu/Cas9 y ARNgu/dCas9, por ejemplo, en una célula o tejido, a diferentes concentraciones. Por ejemplo, los complejos ARNgu/Cas9 y ARNgu/dCas9 se pueden introducir en una proporción de 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1, 3:1 o 2:1. Además, todos estos componentes, es decir, ARNgu, Cas9, dCas9, etc. pueden proporcionarse por separado, por ejemplo, como complejos ensamblados *in vitro* por separado, utilizando constructos de ADN o ARN separados, o juntos, en un solo constructo, o en cualquier combinación. Típicamente, los componentes de ARNgu se complejarán con Cas9 y dCas9 cuando se proporcionen a una célula. Además, las líneas celulares, tal como las células HEK293, pero sin limitarse a ellas, están disponibles comercialmente que expresan constitutivamente Cas9 de *S. pyogenes*, así como fusiones de Cas9-GFP de *S. pyogenes*. En este caso, las células pueden transfectarse sin que Cas9 sea catalíticamente activa, ya que la célula huésped la proporciona.

Los complejos de ARNgu/Cas9 y ARNgu/dCas9 pueden introducirse en diferentes puntos de tiempo. Por ejemplo, los complejos de ARNgu/Cas9 y ARNgu/dCas9 se pueden introducir con al menos 1 minuto de diferencia, 5 minutos de diferencia, 10 minutos de diferencia, 30 minutos de diferencia, 1 hora de diferencia, 5 horas de diferencia, o 15 horas de diferencia o más. Los complejos de ARNgu/Cas9 y ARNgu/dCas9 se pueden introducir como máximo con 1 minuto de diferencia, 5 minutos de diferencia, 10 minutos de diferencia, 30 minutos de diferencia, 1 hora de diferencia, 5 horas de diferencia, o 15 horas de diferencia o más. Los complejos ARNgu/Cas9 se pueden introducir antes que los complejos de ARNgu/dCas9. Los complejos de ARNgu/Cas9 se pueden introducir después de los complejos de ARNgu/dCas9. Los complejos de ARNgu/Cas9 y ARNgu/dCas9 pueden ser regulados diferencialmente (es decir, expresados o estabilizados diferencialmente) a través de agentes suministrados de forma exógena (por ejemplo, promotores de ADN inducibles o proteínas Cas9 inducibles).

Los complejos de ARNgu/Cas9 y ARNgu/dCas9 pueden introducirse en una célula por una variedad de medios que incluyen transfección, transducción, electroporación, micelas y suministro de liposomas, nanopartículas lipídicas, administración viral, inyección de proteínas y similares, que se describen con más detalle más abajo.

Los complejos de ARNgu/dCas9 pueden dirigirse a loci genómicos que se superponen parcialmente. Por ejemplo, estos complejos pueden dirigirse a loci que se superponen por al menos 1, 2, 3, 4, 5, 10, 15, 20, 25, 30 o 35 o más nucleótidos. Estos complejos pueden dirigirse a loci que se superponen como máximo 1, 2, 3, 4, 5, 10, 15, 20, 25, 30 o 35 o más nucleótidos.

Los complejos de ARNgu/dCas9 pueden dirigirse adyacentes a los sitios de actividad nucleasa no deseada observada y la unión de Cas9. Por ejemplo, estos complejos pueden dirigirse a sitios que son adyacentes a un sitio donde se observa actividad no deseada por al menos 1, 2, 3, 4, 5, 10, 15, 20, 25, 30 o 35 o más nucleótidos. Los complejos pueden dirigirse a sitios que son adyacentes a un sitio donde se observa actividad no deseada por un máximo de 1, 2, 3, 4, 5, 10, 15, 20, 25, 30 o 35 o más nucleótidos.

Se pueden usar múltiples complejos de ARN_g/dCas9 para "enlosar" un lugar dado para la actividad de bloqueo de nucleasa máxima. En algunos casos, se utilizan al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más complejos de ARN_g/dCas9. En algunos casos, como máximo se utilizan 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más complejos. Los complejos pueden cubrir un locus. Los complejos pueden cubrir al menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100% de un locus. Los complejos pueden cubrir a lo sumo 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100% de un locus.

Los bloqueadores pueden reducir la unión no deseada de los complejos activos en al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100%. Los bloqueadores pueden reducir la unión no deseada de los complejos activos en un máximo de 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100%.

Sin desear estar ligado a una teoría particular, un complejo de ARN_g/dCas9 puede reducir la unión de un complejo de ARN_g/Cas9 a un ácido nucleico alejado del objetivo por cualquier mecanismo. Por ejemplo, el complejo de ARN_g/dCas9 puede competir con el complejo catalíticamente activo para unirse al ácido nucleico alejado del objetivo. El complejo de ARN_g/dCas9 puede unirse al ácido nucleico alejado del objetivo, creando así un impedimento estérico para el complejo de ARN_g/Cas9 que evita la unión del complejo de ARN_g/Cas9 al ácido nucleico alejado del objetivo. El complejo de ARN_g/dCas9 puede desplazar el complejo de ARN_g/Cas9 del ácido nucleico alejado del objetivo. El complejo de ARN_g/dCas9 puede inhibir que el complejo de ARN_g/Cas9 se una al ácido nucleico alejado del objetivo. El complejo de ARN_g/dCas9 puede bloquear el complejo de ARN_g/Cas9 para que no se una al ácido nucleico alejado del objetivo.

Un complejo de ARN_g/dCas9 puede reducir la unión, la escisión y/o la modificación del ácido nucleico alejado del objetivo por un complejo de ARN_g/Cas9 en un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, o 100%, o cualquier valor dentro de este intervalo. A la inversa, un complejo de ARN_g/dCas9 puede aumentar la unión específica del sitio y/o la modificación en un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o 100% o cualquier valor dentro de este intervalo.

Los procedimientos informáticos para determinar la actividad de la nucleasa no deseada con cualquiera de los procedimientos descritos en este documento pueden comprender escanear la secuencia genómica de un sujeto. La secuencia genómica puede segmentarse *in silico* en una serie de secuencias de ácido nucleico. Las secuencias de ácido nucleico segmentadas pueden alinearse con la secuencia de ácido nucleico dirigida al ácido nucleico. Un algoritmo de búsqueda de secuencia puede determinar una o más secuencias de ácido nucleico no deseadas mediante la identificación de secuencias genómicas segmentadas con alineamientos que comprenden un número definido de pares de bases no emparejadas con el ácido nucleico que se dirige al ácido nucleico. El número de desajustes de pares de bases entre una secuencia genómica y un ácido nucleico que se dirige a un ácido nucleico seleccionado por un algoritmo puede ser definido por el usuario, por ejemplo, el algoritmo puede programarse para identificar secuencias no deseadas con desajustes de hasta cinco pares de bases entre la secuencia genómica y el ácido nucleico que se dirige al ácido nucleico. Los algoritmos de unión *in silico* se pueden utilizar para calcular la eficacia de la unión y/o la escisión de cada secuencia de ácido nucleico alejado del objetivo predicha por un polipéptido dirigido al sitio utilizando un esquema de ponderación. Estos datos se pueden usar para calcular la actividad no deseada para un ácido nucleico dirigido a un ácido nucleico y/o un polipéptido dirigido al sitio.

La actividad de unión no deseada se puede determinar mediante procedimientos experimentales. En un ejemplo no limitante, los procedimientos experimentales pueden comprender la secuenciación de una muestra de ácido nucleico en contacto con un complejo que comprende un polipéptido dirigido al sitio y un ácido nucleico que se dirige al ácido nucleico. La muestra de ácido nucleico contactada se puede fijar o entrecruzar para estabilizar el complejo proteína-ARN-ADN. El complejo que comprende el polipéptido dirigido al sitio, el ácido nucleico (por ejemplo, ácido nucleico objetivo, ácido nucleico alejado del objetivo) y/o el ácido nucleico que se dirige al ácido nucleico puede capturarse de la muestra de ácido nucleico con una etiqueta de afinidad y/o agentes de captura. Se pueden usar técnicas de purificación de ácido nucleico para separar el ácido nucleico objetivo del complejo. Las técnicas de purificación de ácido nucleico pueden incluir la separación en columna de centrifugación, precipitación y electroforesis. El ácido nucleico se puede preparar para el análisis de secuenciación mediante cizallamiento y ligadura de adaptadores. La preparación para el análisis de secuenciación puede incluir la generación de bibliotecas de secuenciación del ácido nucleico objetivo eluido.

Los procedimientos de determinación de secuencia pueden incluir, pero no se limitan a, pirosecuenciación (por ejemplo, como el comercializado por 454 Life Sciences, Inc., Branford, CT); secuenciación por ligadura (por ejemplo, como el comercializado por la tecnología SOLiD^{MR}, Life Technology, Inc., Carlsbad, CA); secuenciación por síntesis usando nucleótidos modificados (como el comercializado por tecnología TruSeq^{MR} y HiSeq^{MR} de Illumina, Inc., San Diego, CA, HeliScope^{MR} de Helicos Biosciences Corporation, Cambridge, MA, y PacBio RS de Pacific Biosciences de California, Inc., Menlo Park, CA); secuenciación por tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, CA); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, CA); tecnologías de secuenciación basadas en nanoporos (por ejemplo, como las desarrolladas por Oxford Nanopore Technologies, LTD, Oxford, Reino Unido), secuenciación capilar (por ejemplo, como la comercializada en la tecnología MegaBACE por Molecular Dynamics, Inc., Sunnyvale, CA), secuenciación electrónica, secuenciación de moléculas individuales (por ejemplo, como la comercializada en tecnología SMRT^{MR} por Pacific Biosciences, Menlo Park, CA), secuenciación de microfluidos en gotas, secuenciación por hibridación (como la comercializada por

Affymetrix, Santa Clara, CA), secuenciación con bisulfito y otros procedimientos de secuenciación conocidos altamente paralelos.

5 En algunos aspectos, la secuenciación se realiza mediante análisis de micromatrices, tal como en el genotipado de SNP por unión. El análisis de secuenciación puede determinar la identidad y la frecuencia de un sitio de unión alejado del objetivo para un ácido nucleico dirigido a un ácido nucleico dado, contando el número de veces que se lee un sitio de unión particular. La biblioteca de ácidos nucleicos secuenciados puede incluir ácidos nucleicos objetivo y ácidos nucleicos alejados del objetivo.

10 La actividad de unión no deseada se puede determinar mediante procedimientos experimentales adicionales. Los procedimientos experimentales pueden comprender la inserción de un oligonucleótido donante en un sitio escindido (Tsai, S.Q. et al., "GUIDE-seq enables genome wide profiling of off-target cleavage by CRISPR-Cas nucleases", Nature Biotech. (2015) 33: 187-197). Luego, el ADN genómico se fragmenta, los adaptadores se agregan y la PCR se realiza con cebadores complementarios al oligonucleótido donante y las secuencias adaptadoras. Las secuencias amplificadas se secuencian y luego se mapean de nuevo con un genoma de referencia. Otros procedimientos experimentales se basan en la explotación de las translocaciones inducidas por ruptura de cadenas dobles del ADN genómico para inducir experimentalmente (a través de la creación de rompimientos de cadena doble) los sitios de "cebo" genómicos (Frock, R.L. et al., "Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases", Nature Biotech. (2015) 33: 179-186). El ADN genómico se fragmenta posteriormente, los adaptadores se agregan y la PCR se realiza con cebadores complementarios al sitio conocido de "cebo" y la secuencia adaptadora. Las secuencias amplificadas se secuencian y luego se mapean de nuevo con un genoma de referencia.

20 En algunos aspectos, las proteínas Cas9 y/o dCas9 pueden modificarse o fusionarse a dominios de proteínas adicionales. Los dominios de proteínas adicionales fusionados pueden mejorar la capacidad de bloquear, alterar o inactivar complejos de Cas9 activos. Los ejemplos de proteínas de fusión que incluyen una proteína Cas9 o dCas9 incluyen, pero no se limitan a, una nucleasa, una transposasa, una metilasa, un represor del factor de transcripción o un dominio activador (por ejemplo, tal como KRAB y VP16), dominios correpresores y coactivadores, ADN metil transferasas, histonas acetiltransferasas, histonas desacetilasas y dominios de escisión de ADN (por ejemplo, un dominio de escisión de la endonucleasa FokI). En algunos aspectos, una secuencia no nativa puede conferir nuevas funciones a la proteína de fusión. Dichas funciones incluyen, pero no se limitan a lo siguiente: actividad de metiltransferasa, actividad de desmetilasa, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad formadora del dímero de pirimidina, actividad de integrasa, actividad de transposasa, actividad de recombinasa, actividad de polimerasa, actividad de ligasa, actividad de helicasa, actividad de fotoliasa, actividad de glicosilasa, actividad de acetiltransferasa, actividad de desacetilasa, actividad de quinasa, actividad de fosfatasa, actividad de ubiquitina ligasa, actividad de desubiquitinación, actividad de adenilación, actividad de desadenilación, actividad de sumoilación, actividad de desumoilación, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación, actividad de remodelación, actividad de proteasa, actividad de oxidoreductasa, actividad de transferasa, actividad de hidrolasa, actividad de liasa, actividad de isomerasa, actividad de sintasa, actividad de sintetasa, actividad de desmiristoilación, y cualquier combinación de los mismos.

25 En algunos casos, se inserta un polinucleótido donante en el ácido nucleico objetivo, cuando el ácido nucleico objetivo se escinde. Los procedimientos pueden, por ejemplo, usarse para modificar el ADN genómico en una célula eucariota aislada de un organismo. Además, los procedimientos también pueden comprender poner en contacto la secuencia objetivo de ácido nucleico en el ADN genómico con un polinucleótido donante en el que la modificación comprende que al menos una parte del polinucleótido donante se integre en la secuencia objetivo de ácido nucleico.

30 La inserción del polinucleótido donante puede realizarse mediante la maquinaria de recombinación homóloga de la célula. El polinucleótido donante puede comprender brazos de homología que son parcial o totalmente complementarios a las regiones del ácido nucleico objetivo fuera del punto de rompimiento. La inserción del polinucleótido donante también puede realizarse mediante una maquinaria de reparación de ADN no homóloga de la célula, en la que no se requieren brazos de homología. Una discusión sobre los polinucleótidos donantes se presenta más detalladamente a continuación.

35 En unos aspectos, el polinucleótido donante puede unirse al complejo de ARN_g/dCas9 para posicionarlo cerca del sitio de escisión dirigido por el complejo de ARN_g/Cas9 activo. Véase, la Fig. 7A. De esta manera, la reparación dirigida por homología, como se describe a continuación, se puede lograr a velocidades más altas.

40 Una realización particular de los procedimientos descritos en este documento se ilustra en las Figs. 4 y 5. La Fig. 4 representa un ejemplo de unión y escisión no deseada de una nucleasa durante la modificación del genoma. Un ácido nucleico **115** objetivo puede ponerse en contacto con un complejo que comprende un polipéptido **105** dirigido al sitio (por ejemplo, Cas9) y un ácido nucleico **110** que se dirige al ácido nucleico (por ejemplo, un ARN_g). El complejo que comprende la Cas9**105** y ARN_g**110** puede unirse a un ácido nucleico **120** objetivo. En algunos casos, el complejo que comprende la Cas9**105** y ARN_g**110** puede unirse a un ácido nucleico **125** alejado del objetivo. En una etapa **130** de escisión, la Cas9 del complejo puede escindir **135** al ácido nucleico **120** objetivo y al ácido nucleico **125** alejado del objetivo, generando así efectos alejados del objetivo.

La Fig. 5 representa una realización a modo de ejemplo de reducción de los eventos de unión y escisión alejados del objetivo utilizando bloqueadores dCas9. Un ácido nucleico **215** objetivo puede ponerse en contacto con un complejo que comprende un polipéptido **205** dirigido al sitio (por ejemplo, Cas9) y un ácido nucleico **210** dirigido al ácido nucleico (por ejemplo, ARN guía único). El complejo que comprende la Cas9**205** y ARNs**210** se puede unir a un ácido nucleico **220** objetivo. En algunos casos, el complejo que comprende la Cas9**205** y ARNs**210** puede unirse a un ácido nucleico **225** alejado del objetivo⁵. Los complejos que comprenden una proteína dCas9 **235** modificada y un ARNs**236** modificado pueden introducirse y ponerse en contacto **230** con el ácido nucleico objetivo. Los complejos dCas9 pueden desplazar o prevenir la unión de complejos que comprenden Cas9**205** activa. La Cas9**205** activa puede escindir **240/245** al ácido nucleico **220** objetivo. La Cas9**205** activa no puede escindir al ácido nucleico **225** alejado del objetivo porque la dCas9**235** está impidiendo su unión y escisión. De esta manera, se puede prevenir la unión y la escisión no deseada.

En otro aspecto, la divulgación está dirigida a un procedimiento para aumentar la eficacia de la inserción de ácido nucleico mediante HDR o mecanismos de reparación no homólogos. Como se explicó anteriormente, múltiples vías de reparación pueden competir en los rompimientos de ADN dirigidos al sitio. Dichos rompimientos pueden repararse, por ejemplo, mediante uniones finales no homólogas (NHEJ) o reparaciones dirigidas por homología (HDR). NHEJ puede dar como resultado una reparación imperfecta y la adición o eliminación de una o más bases, mientras que HDR puede utilizarse para insertar secuencias de ADN exógenas diseñadas racionalmente. La reparación de un rompimiento de cadena doble (DSB) en presencia de un polinucleótido donante da como resultado una porción de rompimientos reparados fielmente por HDR y se acopla una porción de rompimientos en los que se implementa otra vía de reparación menos confiable, tal como NHEJ, lo que resulta en resultados de reparación mixtos. Las vías alternativas de reparación para la inserción de ADN usando mecanismos no homólogos también pueden resultar en la inserción de ADN donante en el sitio de rompimiento.

HDR se basa en la presencia de un polinucleótido donante, un fragmento de ADN que comparte homología con secuencias en o cerca de un rompimiento de ADN que se puede usar para reparar rompimientos de ADN. Sin desear estar limitado por ninguna teoría o mecanismo particular, en algunos aspectos, la presente invención proporciona procedimientos para usar polipéptidos dirigidos al sitio (por ejemplo, nucleasas Cas9) para crear un sustrato que se acople a una ruta HDR alternativa, similar a la rama de hibridación de cadena sencilla (SSA) de HDR, y evitará la competición de las rutas de reparación del ADN, tales como NHEJ, de reparación del rompimiento.

La hibridación de cadena sencilla (SSA) es un proceso que se inicia cuando se introduce un rompimiento entre dos secuencias repetitivas orientadas en la misma dirección. En general, son necesarios cuatro etapas para la reparación de los rompimientos por SSA: (1) una etapa de resección final que extiende las secuencias repetidas y forma el ADN de cadena sencilla 3' largo; (2) una etapa de hibridación en la que las dos secuencias repetitivas se hibridan juntas formando una estructura de solapa; (3) una segunda etapa de resección en la que las estructuras de solapa formadas por las regiones entre las repeticiones se resecan y; (4) ligación de los extremos. HDR en las muescas de ADN se produce a través de un mecanismo a veces denominado "HDR alternativo" que comparte muchas de las mismas dependencias genéticas de la SSA, tal como la inhibición por RAD51 y BRCA2.

Los inventores de la presente invención han desarrollado un sistema CRISPR modificado generando al menos dos muescas de cadena sencilla en la misma cadena de un ácido nucleico de cadena doble objetivo y proporcionando un polinucleótido donante que puede hibridarse con la cadena sin muesca. Esto da como resultado la inserción precisa de ADN exógeno con poca unión final mutagénica de fondo.

Este procedimiento emplea moléculas de Cas9 en tándem que comprenden una o más mutaciones que convierten las moléculas de Cas9 catalíticamente activas en nickasas. Las nickasas se dirigen a sitios específicos utilizando ARNs diseñados para dirigirse a dos sitios en la misma cadena en un ácido nucleico objetivo de cadena doble, para generar dos muescas (es decir, rompimientos de cadena sencilla) en la cadena seleccionada.

Se puede usar cualquier molécula de Cas9, como se describe en detalle más arriba, siempre que la Cas9 funcione como una nickasa. En algunos aspectos, esto se puede lograr mediante la introducción de una mutación puntual en cualquiera de los dos residuos catalíticos (D10A y H840A, numerados en relación con Cas9 de *S. pyogenes*) del gen que codifica Cas9. Una mutación de aminoácidos en cualquiera de las posiciones en Cas9 da como resultado la inactivación de la actividad catalítica de la nucleasa y convierte a Cas9 en una enzima nickasa que produce rompimientos de cadena sencilla en los sitios objetivo. La Cas9 doble mutante con cambios en las posiciones de aminoácidos D10A y H840A, sin embargo, inactiva completamente las actividades de nucleasa y nickasa. La especificidad de direccionamiento se determina mediante el emparejamiento de bases complementarias de un ARN con los loci genómicos que incluyen secuencias de PAM adyacentes a las mismas.

Las nickasas pueden comprender cualquier mutación que permita a Cas9 escindir solo una cadena de un ácido nucleico objetivo de cadena doble. Por ejemplo, como se explicó anteriormente, la Cas9 (por ejemplo, Cas9 de *S. pyogenes*) puede comprender una mutación D10A en uno de sus dominios de nucleasa, o en un residuo correspondiente en una Cas9 ortóloga para volver la molécula una nickasa. La Cas9 (por ejemplo, Cas9 de *S. pyogenes*) puede comprender una mutación H840A en uno de sus dominios de nucleasa, o un residuo correspondiente en una Cas9 ortóloga para volver la molécula una nickasa.

Por consiguiente, se puede usar cualquier molécula de Cas9 que tenga actividad de nickasa y solo produzca rompimientos de cadena sencilla. Como se explicó anteriormente, las proteínas Cas9 son conocidas y las proteínas Cas9 pueden derivarse de cualquier especie bacteriana, subespecie o cepa que codifique la misma. También son útiles las proteínas codificadas por los ortólogos de Cas9, las proteínas sintéticas de tipo Cas9 y sus variantes y modificaciones. En un aspecto, los ortólogos de Cas9 se seleccionan basándose en las secuencias de motivos adyacentes protoespaciadores (PAM) particulares presentes en el ácido nucleico objetivo. Por ejemplo, Cas9 de *S. pyogenes* apunta a secuencias NGG. Un experto en la técnica puede determinar fácilmente la Cas9 particular a mutar con base en la especificidad particular deseada.

Además, las nickasas utilizadas en los presentes procedimientos deben emparejarse de modo que las muescas se produzcan en la misma cadena. Por ejemplo, ambas nickasas utilizadas pueden incluir una mutación D10A, o ambas pueden incluir una mutación H840A. Un nickasa puede ser una nickasa Cas9 de *S. pyogenes* y la otra puede ser una nickasa que apunta a un PAM con una secuencia adyacente diferente a la seleccionada por la nickasa Cas9 de *S. pyogenes*, tal como una nickasa diseñada a partir de una proteína Cas9 ortóloga, siempre que la misma cadena esté mellada. Las nickasas apropiadas para uso en los presentes procedimientos se basan por lo tanto en la secuencia objetivo de ácido nucleico y en una determinación de secuencias adyacentes a PAM presentes en los sitios de escisión deseados. De esta manera, el procedimiento proporciona flexibilidad para la escisión de cadena sencilla del ácido nucleico objetivo.

Las nickasas pueden escindir la cadena sentido del ácido nucleico objetivo de cadena doble o la cadena antisentido del ácido nucleico objetivo de cadena doble (por ejemplo, ADN). Las nickasas pueden escindir la misma cadena del ácido nucleico objetivo de cadena doble.

Las dos nickasas pueden diseñarse para escindir a una distancia de al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100, 500, 1.000 o 5.000 o más bases separadas unas de otras. Las dos nickasas pueden diseñarse para escindir a una distancia de a lo sumo 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100, 500, 1.000 o 5.000 o más bases separadas unas de otras. La distancia entre las muescas determinará la longitud del polinucleótido donante que se proporcionará para la inserción.

Como se explicó anteriormente, una vez que el ácido nucleico objetivo se corta, un polinucleótido donante puede dirigirse e insertarse en un sitio objetivo de interés para modificar el ácido nucleico objetivo. Las modificaciones de ADN seleccionadas que utilizan polinucleótidos donantes para grandes cambios (por ejemplo, inserciones o supresiones de más de 100 pb) utilizan tradicionalmente plantillas de donantes basadas en plásmidos que contienen brazos de homología que flanquean el sitio de la alteración. Cada brazo puede variar en longitud, pero típicamente es más largo que aproximadamente 200 pb para una inserción grande, el tamaño de los brazos depende del tamaño del polinucleótido donante y el polinucleótido objetivo, como se explica en detalle más arriba.

Para modificaciones más cortas (por ejemplo, menos de 50 pb), se pueden usar oligonucleótidos de cadena sencillas tales como oligonucleótidos de ADN, oligonucleótidos parcialmente de cadena doble, donantes de cadena doble cortados, y similares, en lugar de plásmidos de direccionamiento. En este aspecto, por ejemplo, se pueden usar oligonucleótidos de cadena sencilla que contienen secuencias flanqueantes con homología en la proximidad de cada corte, y se pueden orientar en dirección sentido o antisentido en relación con el locus objetivo. Para el diseño de oligonucleótidos de ADN de cadena sencilla, se usa típicamente un oligonucleótido con una homología total de alrededor de 100-150 pb. La mutación se introduce en el medio, produciendo brazos de homología de aproximadamente 50-75 pb. Sin embargo, estos números pueden variar, dependiendo del tamaño del polinucleótido donante y el polinucleótido objetivo. También se pueden usar donantes de polinucleótidos no simétricos que están compuestos por dos oligonucleótidos de ADN. Son parcialmente complementarios, y cada uno incluye una región flanqueante de homología. Para algunas modificaciones, el polinucleótido donante puede tener al menos un brazo con aproximadamente 10 bases de homología con la secuencia objetivo. Para algunas modificaciones, el polinucleótido donante puede tener al menos un brazo con menos de 100 bases de homología con la secuencia objetivo. Para otras modificaciones, el donante puede tener más de 100 bases de homología con la secuencia objetivo. En algunos casos, el donante puede tener brazos de homología de la misma longitud. En otros casos, el donante puede tener brazos de homología de diferentes longitudes. En algunos casos, al menos uno de los brazos de homología es de longitud cero.

Por lo tanto, un polinucleótido donante se puede diseñar para hibridar con el espacio de cadena sencilla que resulta de los cortes hechos por las dos nickasas. Como se explicó anteriormente, el polinucleótido donante puede comprender adicionalmente regiones de homología con las secuencias fuera de los rompimientos. El tamaño de las regiones de homología se determinará por el tamaño del polinucleótido objetivo y puede ser de al menos 5, 10, 15, 20, 25, 30, 35 o más nucleótidos de longitud, el tamaño depende del tamaño del polinucleótido donante y el ácido nucleico objetivo. Las regiones de homología pueden tener como máximo 5, 10, 15, 20, 25, 30, 35 o más nucleótidos de longitud. El polinucleótido donante puede ser de cadena sencilla. El polinucleótido donante de cadena sencilla se puede insertar en el rompimiento creado por las dos nickasas en tándem.

La Fig. 6 representa un aspecto a modo de ejemplo de los presentes procedimientos. En este caso, se utilizan dos nickasas Cas9 D10A en tándem para cortar una región de cadena sencilla de ADN en la misma cadena de un ácido nucleico de cadena doble objetivo. Como se muestra en la Fig. 6A, dos nickasas Cas9 (en este caso, las nickasas

Cas9 de *S. pyogenes* con mutaciones de D10A en el dominio de la endonucleasa HNH) se dirigen a dos posiciones separadas en la cadena sentido de un polinucleótido objetivo utilizando dos complejos de ARNg/nickasa Cas9. El direccionamiento se logra utilizando una secuencia espaciadora presente en el ARNg que se ha diseñado para apuntar específicamente a una región complementaria en la secuencia de ácido nucleico objetivo. La especificidad de unión se determina mediante el emparejamiento de bases de ARNg-ADN y el PAM, en este caso, NGG, yuxtapuesto a la región complementaria de ADN (véase, por ejemplo, Mojica F.J. et al., "Short motif sequences determine the targets of the prokaryotic CRISPR defence system" *Microbiology* (2009) 155: 733-740; Shah S.A. et al., "Protospacer recognition motifs: mixed identities and functional diversity" *RNA Biology* (2013) 10: 891-899; Jinek M. et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity" *Science* (2012) 337: 816-821). La cadena sencilla seleccionada se escinde (Fig. 6B) y el donante, con regiones flanqueantes superpuestas, se inserta (Fig. 6C).

En otro aspecto, la divulgación está dirigida a procedimientos adicionales para aumentar la HDR. La metodología actual para introducir un cambio deseado en un gen incluye transfectar, electroporar o microinyectar una endonucleasa específica del sitio y moléculas donantes en una célula o embrión y usar la difusión pasiva para ubicar las moléculas donantes en todo el núcleo ((Lin, S. et al. "Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery", *eLife* (2014) Diciembre; doi: 10.7554 / eLife.04766). Sin embargo, este procedimiento de HDR típicamente tiene baja eficiencia. A diferencia de la difusión pasiva, los procedimientos que se describen a continuación colocan la molécula donante cerca del sitio de corte para aumentar la eficiencia de la HDR.

En estos procedimientos, se usan uno o más complejos de ARNg/dCas9, junto con un complejo de ARNg/Cas9 catalíticamente activo. Los uno o más complejos de ARNg/dCas9 incluyen un donante polinucleótido asociado con el mismo para posicionar el polinucleótido donante cerca de un sitio objetivo con el fin de aumentar la eficacia de HDR. Por lo tanto, la dCas9 atada puede posicionar la molécula donante en una orientación que aumentará la probabilidad de que la molécula donante se incorpore en el sitio objetivo a través de HDR, introduciendo así un cambio deseado en la secuencia objetivo.

Como se explicó anteriormente, el polinucleótido donante puede ser un ADN, ARN, un vector, un plásmido o similar de cadena sencilla o doble, y puede usarse para transferir información genética (es decir, secuencias de polinucleótidos) del donante en el sitio del rompimiento en el ácido nucleico objetivo. El donante se puede usar para insertar o reemplazar secuencias polinucleotídicas en una secuencia objetivo, por ejemplo, para introducir un polinucleótido que codifica una proteína o ARN funcional (por ejemplo, ARNip), para introducir una etiqueta de proteína, para modificar una secuencia reguladora de un gen, o para introducir una secuencia reguladora en un gen (por ejemplo, un promotor, un potenciador, una secuencia de entrada al ribosoma interno, un codón de inicio, un codón de parada, una señal de localización o una señal de poliadenilación), para modificar una secuencia de ácido nucleico (por ejemplo, introducir una mutación), y similares.

Se puede usar un solo complejo de ARNg/dCas9 con el donante asociado, como se muestra en la Fig. 7A. Alternativamente, dos de estos complejos se pueden usar para colocar al donante a través del sitio de corte como se muestra en la Fig. 7B. Las moléculas dCas9 y Cas9 y los polinucleótidos guía utilizados en los complejos pueden ser cualquiera de los descritos anteriormente.

Cuando se usa un complejo de ARNg/dCas9, el complejo puede dirigirse al ácido nucleico en sentido ascendente o descendente del ácido nucleico seleccionado por el complejo de ARNg/Cas9 catalíticamente activo. Un polinucleótido donante está asociado con el complejo de ARNg/dCas9. De esta manera, el polinucleótido donante se pone en proximidad con el ácido nucleico objetivo escindido y la HDR insertará al menos una parte del polinucleótido donante en el sitio escindido.

Cuando se usan dos complejos de ARNg/dCas9, el segundo complejo de ARNg/dCas9 está diseñado para dirigirse al ácido nucleico secuencia abajo del complejo de ARNg/Cas9 catalíticamente activo cuando el primer ARNg/dCas9 dirige el ácido nucleico secuencia arriba del complejo de ARNg/Cas9 catalíticamente activo. Alternativamente, el segundo complejo de ARNg/dCas9 está diseñado para dirigirse al ácido nucleico secuencia arriba del complejo de ARNg/Cas9 catalíticamente activo cuando el primer ARNg/dCas9 se dirige al ácido nucleico secuencia abajo del complejo de ARNg/Cas9 catalíticamente activo. Por lo tanto, el objetivo para el complejo de ARNg/Cas9 activo está en una posición entre los dos complejos inactivos. Además, el extremo 5' del donante polinucleótido se asociará con uno de los complejos inactivos de ARNg/dCas9 y el extremo 3' asociado con el otro de los complejos inactivos, de modo que el donante polinucleótido se posicione a través del sitio de escisión para la inserción utilizando HDR. Un experto en la materia puede determinar fácilmente qué extremo del polinucleótido donante se asocia con cada complejo basándose en el objetivo deseado.

El donante está unido a los complejos utilizando procedimientos bien conocidos en la técnica. Para hacerlo, la cadena principal del ARNg puede extenderse para incluir una región complementaria a la molécula donante. Por ejemplo, el ARNg en el complejo de ARNg/dCas9 puede incluir varios nucleótidos adicionales, por ejemplo, 5-20, tal como 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o incluso más, nucleótidos adicionales en el extremo 3' del ARNg que se unirán de manera complementaria al extremo 5' o 3' de un polinucleótido donante de ADN de cadena sencilla. De esta manera, el polinucleótido donante se posicionará para interactuar con el sitio de corte

inducido por ARN_g/Cas9 y la maquinaria de HDR endógena de la célula incorporará al donante en el sitio de escisión. El polinucleótido donante anclado ARN_g/dCas9 se coloca secuencia arriba o secuencia abajo del rompimiento de cadena doble y está disponible a una mayor concentración local para HDR.

5 En todas las realizaciones de los procedimientos descritos anteriormente, los diversos componentes pueden proporcionarse a una célula o *in vitro*, por ejemplo, usando casetes de expresión que codifican una Cas9, una dCas9, ARN_g; un polinucleótido donante, etc. Estos componentes pueden estar presentes en un solo casete o casetes múltiples, en la misma o diferentes constructos. Los casetes de expresión típicamente comprenden secuencias reguladoras que están involucradas en uno o más de los siguientes: regulación de la transcripción, regulación postranscripcional y regulación de la traducción. Los casetes de expresión pueden introducirse en una amplia variedad de organismos, incluyendo células bacterianas, células de levadura, células vegetales y células de mamíferos. Los casetes de expresión comprenden típicamente secuencias reguladoras funcionales correspondientes al organismo u organismos en los que se están introduciendo.

15 En un aspecto, todos o una parte de los diversos componentes de los procedimientos se proporcionan en vectores, incluidos vectores de expresión, que comprenden polinucleótidos que codifican una Cas9, una dCas9, un ARN_g y/o un polinucleótido donante. Los vectores útiles para poner en práctica la presente invención incluyen plásmidos, virus (incluidos los fagos) y fragmentos de ADN de la tabla de Integra (es decir, fragmentos integrables en el genoma del huésped mediante recombinación homóloga). Un vector se replica y funciona independientemente del genoma huésped, o puede, en algunos casos, integrarse en el mismo genoma. Los vectores de replicación adecuados contendrán un replicón y secuencias de control derivadas de especies compatibles con la célula huésped de expresión deseada. Las células huésped transformadas son células que se han transformado o transfectado con los vectores construidos utilizando técnicas de ADN recombinante.

25 Los procedimientos generales para la construcción de vectores de expresión son conocidos en la técnica. Los vectores de expresión para la mayoría de las células huésped están disponibles comercialmente. Existen varios productos de software comerciales diseñados para facilitar la selección de vectores apropiados y su construcción, tal como los vectores de células de insectos para la transformación de células de insectos y la expresión de genes en células de insectos, los plásmidos bacterianos para la transformación bacteriana y la expresión de genes en células bacterianas, los plásmidos de levadura para la transformación de células y la expresión de genes en levaduras y otros hongos, vectores de mamíferos para la transformación de células de mamíferos y la expresión de genes en células de mamíferos o mamíferos, vectores virales (incluidos los vectores retrovirales, lentivirales y adenovirales) para la transformación de células y la expresión de genes y procedimientos para permitir fácilmente la clonación de tales polinucleótidos. SnapGene^{MR} (GSL Biotech LLC, Chicago, Illinois; snapgene.com/resources/plasmid_files/your_time_is_val_valuable/), por ejemplo, proporciona una extensa lista de vectores, secuencias de vectores individuales y mapas de vectores, así como fuentes comerciales para muchos de los vectores.

35 Los casetes de expresión típicamente comprenden secuencias reguladoras que están involucradas en uno o más de los siguientes: regulación de la transcripción, regulación postranscripcional y regulación de la traducción. Los casetes de expresión pueden introducirse en una amplia variedad de organismos, incluidas las células bacterianas, células de levadura, células de mamíferos y células vegetales. Los casetes de expresión comprenden típicamente secuencias reguladoras funcionales correspondientes a las células huésped u organismo u organismos en los que se introducen. Los vectores de expresión también pueden incluir polinucleótidos que codifican etiquetas de proteínas (por ejemplo, etiquetas de poli-His, etiquetas de hemaglutinina, etiquetas de proteínas fluorescentes, etiquetas bioluminiscentes, etiquetas de localización nuclear). Las secuencias de codificación para tales etiquetas de proteínas pueden fusionarse con las secuencias de codificación o pueden incluirse en un casete de expresión, por ejemplo, en un vector de direccionamiento.

45 En algunos aspectos, los polinucleótidos que codifican uno o más de los diversos componentes están unidos operativamente a un promotor inducible, un promotor reprimible o un promotor constitutivo.

50 Se han diseñado varios vectores de expresión para expresar polinucleótidos guía. Véase, por ejemplo, Shen, B. et al. "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects" (2014), 2 de marzo. Doi: 10.1038/nmeth.2857.10.1038/nmeth.2857. Además, los vectores y los sistemas de expresión están disponibles comercialmente, tales como a través de New England Biolabs (Ipswich, MA) y Clontech Laboratories (Mountain View, CA). Los vectores pueden diseñarse para expresar simultáneamente un ARN_g específico del objetivo utilizando un promotor U2 o U6, una Cas9 y/o dCas9, y si se desea, una proteína marcadora, para monitorear la eficiencia de la transfección y/o para enriquecer/aislar más las células transfectadas mediante citometría de flujo.

55 Los vectores pueden diseñarse para la expresión de diversos componentes de los procedimientos descritos en células procariontas o eucariotas. Alternativamente, la transcripción puede ser *in vitro*, por ejemplo, utilizando secuencias reguladoras del promotor T7 y polimerasa T7. Se pueden usar otras secuencias de ARN polimerasa y promotoras.

Los vectores pueden introducirse y propagarse en un procarionta. Los vectores procariontas son bien conocidos en la técnica. Típicamente, un vector procarionta comprende un origen de replicación adecuado para la célula huésped

objetivo (por ejemplo, oriC derivado de *E. coli*, pUC derivado de pBR322, pSC101 derivado de *Salmonella*), origen 15A (derivado de p15A) y cromosomas artificiales bacterianos). Los vectores pueden incluir un marcador seleccionable (por ejemplo, genes que codifican resistencia para ampicilina, cloranfenicol, gentamicina y kanamicina). Zeocin^{MR} (Life Technologies, Grand Island, NY) se puede utilizar como una selección en bacterias, hongos (incluida la levadura), plantas y líneas celulares de mamíferos. En consecuencia, se pueden diseñar vectores que contengan solo un gen de resistencia a fármacos para la Zeocina para el trabajo de selección en varios organismos. Se conocen promotores útiles para la expresión de proteínas en procariotas, por ejemplo, T5, T7, ramnosa (inducible), arabinosa (inducible) y PhoA (inducible). Además, los promotores T7 se usan ampliamente en vectores que también codifican la ARN polimerasa T7. Los vectores procariotas también pueden incluir sitios de unión a ribosomas de fuerza variable y señales de secreción (por ejemplo, mal, sec, tat, ompC y pelB). Además, los vectores pueden comprender promotores de ARN polimerasa para la expresión de ARN_g. Las secuencias de terminación de la transcripción de la ARN polimerasa procariota también son bien conocidas (por ejemplo, las secuencias de terminación de la transcripción de *S. pyogenes*).

Los vectores de integración para la transformación estable de procariotas también se conocen en la técnica (véase, por ejemplo, Heap, J. T., et al., "Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker", *Nucleic Acids Res.* (2012) 40: e59).

La expresión de proteínas en procariotas se lleva a cabo típicamente en *Escherichia coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o no de fusión.

Una amplia variedad de promotores de ARN polimerasa adecuados para la expresión de los diversos componentes están disponibles en procariotas (véase, por ejemplo, Jiang, Y., et al., "Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system," *Environ Microbiol.* (2015) 81: 2506-2514); Estrem, S.T., et al., (1999) "Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit", *Genes Dev.* 15;13(16): 2134-47).

En algunos aspectos, un vector es un vector de expresión de levadura que comprende uno o más componentes de los procedimientos descritos anteriormente. Los ejemplos de vectores para la expresión en *Saccharomyces cerevisiae* incluyen, pero no se limitan a, lo siguiente: pYepSec1, pMFa, pJRY88, pYES2, y picZ. Los procedimientos para la expresión génica en células de levadura son conocidos en la técnica (véase, por ejemplo, *Methods in Enzymology*, Volume 194, "Guide to Yeast Genetics and Molecular and Cell Biology, Parte A", (2004) Christine Guthrie y Gerald R. Fink (eds.), Elsevier Academic Press, San Diego, CA). Típicamente, la expresión de genes que codifican proteínas en levaduras requiere un promotor operativamente unido a una región codificante de interés más un terminador transcripcional. Se pueden usar diversos promotores de levadura para construir casetes de expresión para expresión de genes en levadura. Los ejemplos de promotores incluyen, pero no se limitan a, promotores de genes que codifican las siguientes proteínas de levadura: gen de la alcohol deshidrogenasa 1 (ADH1) o alcohol deshidrogenasa 2 (ADH2), fosfoglicerato quinasa (PGK), triosa fosfato isomerasa (TPI), gliceraldehído-3-fosfato deshidrogenasa (GAPDH; también conocida como TDH3, o triosa fosfato deshidrogenasa), galactosa-1-fosfato uridil-transferasa (GAL7), UDP-galactosa epimerasa (GAL10), citocromo c1 (CYC1), ácido fosfatasa (PHO5) y glicerol-3-fosfato deshidrogenasa (GPD1). También se pueden usar promotores híbridos, tales como los promotores ADH2/GAPDH, CYC1/GAL10 y ADH2/GAPDH (que se induce a bajas concentraciones de glucosa celular, por ejemplo, aproximadamente 0,1 por ciento a aproximadamente 0,2 por ciento). En *S. pombe*, los promotores adecuados incluyen al promotor nmt1 reprimido con tiamina y el promotor de citomegalovirus constitutivo en pTL2M.

Los promotores de ARN polimerasa III de levadura (por ejemplo, promotores de los genes 5S, U6 o RPR1), así como las secuencias de terminación de la polimerasa III, son conocidos en la técnica (véase, por ejemplo, www.yeastgenome.org; Harismendy, O., et al., (2003) "Genome-wide location of yeast RNA polymerase III transcription machinery" *The EMBO Journal.* 22(18): 4738-4747).

Además de un promotor, se pueden usar varias secuencias de activación secuencia arriba (UAS), también llamadas potenciadores, para mejorar la expresión del polipéptido. Los ejemplos de secuencias de activación secuencia arriba para la expresión en levadura incluyen las UAS de genes que codifican estas proteínas: CYC1, ADH2, GAL1, GAL7, GAL10 y ADH2. Los ejemplos de secuencias de terminación de la transcripción para la expresión en levaduras incluyen las secuencias de terminación de los genes factor α , CYC1, GAPDH y PGK. Se pueden usar una o múltiples secuencias de terminación.

Los promotores, terminadores y regiones de codificación adecuados pueden clonarse en vectores lanzadera en *E. coli* y transformados en células de levadura. Estos vectores permiten la propagación de la cepa tanto en cepas de levadura como de *E. coli*. Normalmente, el vector contiene un marcador y secuencias seleccionables que permiten la replicación autónoma o la integración cromosómica en cada huésped. Ejemplos de plásmidos utilizados típicamente en levaduras son los vectores lanzadera pRS423, pRS424, pRS425 y pRS426 (American Type Culture Collection, Manassas, VA). Estos plásmidos contienen un origen de replicación de 2 micras de levadura, un origen de replicación de *E. coli* (por ejemplo, pMB1) y un marcador seleccionable.

Los diversos componentes también pueden expresarse en insectos o células de insecto. Las secuencias de control de expresión adecuadas para uso en tales células son bien conocidas en la técnica. En algunos aspectos, es

- deseable que la secuencia de control de expresión comprenda un promotor constitutivo. Los ejemplos de promotores fuertes adecuados incluyen, entre otros, los siguientes: los promotores de baculovirus para los genes *piO*, de polihedrina (*polh*), *p 6.9*, cápside, UAS (contiene un sitio de unión a Gal4), *Ac5*, tipo cathepsina, promotor del gen de actina de *B mori*; promotores del gen *hsp70* de *Drosophila melanogaster*, actina, α -1 tubulina o ubiquitina, promotores de RSV o MMTV, promotor copia, promotor gitano y promotor del gen IE de citomegalovirus. Los ejemplos de promotores débiles que pueden usarse incluyen, pero no se limitan a, los siguientes: los promotores de baculovirus para los genes *iel*, *ie2*, *ieO*, *etl*, *39K* (también conocido como *pp31*) y *gp64*. Si se desea aumentar la cantidad de expresión génica a partir de un promotor débil, se pueden usar elementos potenciadores, tales como el elemento potenciador de baculovirus, *hr5*, junto con el promotor.
- Para la expresión de algunos de los componentes divulgados en el presente documento en insectos, los promotores de ARN polimerasa III son conocidos en la técnica, por ejemplo, el promotor U6. También se conocen las características conservadas de los promotores de ARN polimerasa III en insectos (véase, por ejemplo, Hernandez, G., (2007) "Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity", *Nucleic Acids Res.* Enero de 2007; 35(1): 21-34).
- En otro aspecto, los diversos componentes se incorporan en vectores de mamíferos para uso en células de mamíferos. Un gran número de vectores de mamífero adecuados para uso con los sistemas de la presente invención están disponibles comercialmente (por ejemplo, a través de Life Technologies, Grand Island, NY; NeoBiolab, Cambridge, MA; Promega, Madison, WI; DNA2.0, Menlo Park, CA; Addgene, Cambridge, MA).
- Los vectores derivados de virus de mamífero también se pueden usar para expresar los diversos componentes de los presentes procedimientos en células de mamíferos. Estos incluyen vectores derivados de virus tales como adenovirus, papovirus, herpesvirus, poliomavirus, citomegalovirus, lentivirus, retrovirus, vacuna y virus 40 de simio (SV40) (véase, por ejemplo, Kaufman, R. J., (2000) "Overview of vector design for mammalian gene expression", *Molecular Biotechnology*, Volumen 16, Número 2, páginas 151-160; Cooray S., et al., (2012) "Retrovirus and lentivirus vector design and methods of cell conditioning", *Methods Enzymol.* 507: 29-57). Las secuencias reguladoras operativamente vinculadas a los componentes pueden incluir secuencias de unión activadoras, potenciadores, intrones, secuencias de reconocimiento de poliadenilación, promotores, secuencias de unión a represor, estructuras de tallo-bucle, secuencias de inicio translacionales, secuencias líder de la traducción, secuencias de terminación de la transcripción, secuencias de terminación de la traducción, sitios de unión del cebador, y similares. Los promotores de uso común son promotores constitutivos de mamíferos CMV, EF1a, SV40, PGK1 (ratón o humano), Ubc, CAG, CaMKIIa y beta-Act y otros conocidos en la técnica (Khan, K. H. (2013) "Gene Expression in Mammalian Cells and its Applications", *Advanced Pharmaceutical Bulletin* 3(2), 257-263). Además, pueden usarse promotores de ARN polimerasa III de mamíferos, que incluyen HI y U6.
- En algunos aspectos, un vector de expresión recombinante de mamífero es capaz de dirigir preferentemente la expresión del ácido nucleico en un tipo de célula particular (por ejemplo, usando elementos reguladores específicos de tejido para expresar un polinucleótido). Los elementos reguladores específicos de tejido son conocidos en la técnica e incluyen, entre otros, el promotor de albúmina, los promotores específicos de linfoides, los promotores específicos de neuronas (por ejemplo, el promotor de neurofilamento), los promotores específicos de páncreas y los promotores específicos de glándulas mamarias (por ejemplo, promotor de suero de leche), y en particular promotores de receptores de células T e inmunoglobulinas. Los promotores regulados por desarrollo también están incluidos, por ejemplo, los promotores *hox* murinos y el promotor de alfafetoproteína.
- Se han utilizado numerosas líneas celulares de mamíferos para la expresión de productos génicos que incluyen HEK 293 (riñón embrionario humano) y CHO (ovario de hámster chino). Estas líneas celulares pueden transfectarse mediante procedimientos estándar (por ejemplo, utilizando fosfato de calcio o polietilimina (PEI) o electroporación). Otras líneas celulares típicas de mamíferos incluyen, pero no se limitan a: HeLa, U2OS, 549, HT1080, CAD, P19, NIH 3T3, L929, N2a, células 293 de riñón embrionario humano, MCF-7, Y79, SO-Rb50, Hep G2, DUKX-X11, J558L y células de riñón de hámster bebe (BHK).
- Los procedimientos para introducir polinucleótidos (por ejemplo, un vector de expresión) en células huésped son conocidos en la técnica y se seleccionan típicamente en función del tipo de célula huésped. Tales procedimientos incluyen, por ejemplo, infección viral o bacteriófaga, transfección, conjugación, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilimina, transfección mediada por DEAE-dextrano, fusión de protoplastos, lipofección, transfección mediada por liposomas, tecnología de pistola de partículas, microinyección directa, y la liberación mediada por nanopartículas.
- Como se explicó anteriormente, un aspecto de la presente invención proporciona procedimientos para aumentar la eficiencia de la modificación del genoma mediada por Cas9 ya sea disminuyendo el número de rompimientos de cadena doble del ácido nucleico alejado del objetivo, y/o mejorando la HDR de un ácido nucleico objetivo escindido, modificando así los genomas utilizando HDR. La presente divulgación también incluye procedimientos de modulación de transcripción *in vitro* o *in vivo* usando los diversos componentes y complejos descritos en el presente documento. En un aspecto, un complejo de ARN*gu*/proteína dCas puede reprimir la expresión génica al interferir con la transcripción cuando el ARN*gu* dirige la unión diana ADN del complejo a la región promotora del gen. El uso de los complejos para reducir la transcripción también incluye complejos en los que la proteína dCas se fusiona con un

regulador descendente conocido de un gen objetivo (por ejemplo, un polipéptido represor). Por ejemplo, la expresión de un gen está bajo el control de secuencias reguladoras a las que se puede unir un polipéptido represor. Un polinucleótido guía puede dirigir la unión diana ADN de un complejo de proteína represora a las secuencias de ADN que codifican las secuencias reguladoras o adyacentes a las secuencias reguladoras, de modo que la unión del complejo de proteína represora pone a la proteína represora en contacto operable con las secuencias reguladoras. De manera similar, dCas9 puede fusionarse con un polipéptido activador para activar o aumentar la expresión de un gen bajo el control de secuencias reguladoras a las que se puede unir un polipéptido activador.

Otro aspecto es el uso de complejos de ARN_g/dCas9 en procedimientos para aislar o purificar regiones de ADN genómico (ADN_g). En un aspecto del procedimiento, una proteína dCas se fusiona con un epítipo (por ejemplo, un epítipo FLAG®, Sigma Aldrich, St. Louis, MO) y un ARN_g dirige la unión del ADN diana de un complejo de ARN_g/proteína dCas9-epítipo a secuencias de ADN dentro de la región de ADN genómico que se va a aislar o purificar. Se utiliza un agente de afinidad para unir el epítipo y el ADN_g asociado al complejo de ARN_g/proteína dCas9-epítipo.

La presente divulgación también abarca procedimientos de terapia génica para prevenir o tratar enfermedades, trastornos y afecciones usando los diversos procedimientos descritos en el presente documento. En un aspecto, un procedimiento de terapia génica usa la introducción de secuencias de ácido nucleico en un organismo o células de un organismo (por ejemplo, un paciente) para lograr la expresión de los componentes de la presente invención para proporcionar la modificación de una función objetivo. Por ejemplo, las células de un organismo pueden modificarse, *ex vivo*, mediante (i) introducción de vectores que comprenden casetes de expresión que expresan los diversos componentes, (ii) introducción directa de ARN_g y/o polinucleótidos donantes y proteínas Cas9 y/o dCas9, o (iii) introducción de combinaciones de estos componentes. Las células modificadas se proporcionan a un organismo (por ejemplo, un paciente) para ser tratado.

En la técnica se conocen ejemplos de técnicas de terapia génica y administración para la terapia (véase, por ejemplo, Kay, M.A., (2011) "State-of-the-art gene-based therapies: the road ahead," *Nature Reviews Genetics* 12, 316-328; Wang, D., et al., (2014) "State-of-the-art human gene therapy: parte I. Gene delivery technologies," *Discov Med.* 18(97): 67-77; Wang, D., et al., (2014) "State-of-the-art human gene therapy: parte II. Gene therapy strategies and clinical applications," *Discov Med.* 18(98): 151-61; "The Clinibook: Clinical Gene Transfer State of the Art," Odile Cohen-Haguener (Editor), EDP Sciences (Octubre 31, 2012), ISBN-10: 2842541715).

En algunos aspectos, los componentes divulgados en el presente documento se suministran utilizando sistemas de administración a nanoescala, tales como nanopartículas. Además, se pueden usar liposomas y otros sistemas de suministro de partículas. Por ejemplo, los vectores que incluyen los componentes de los presentes procedimientos pueden envasarse en liposomas antes de su entrega al sujeto o a las células derivadas de los mismos, tal como se describe en las patentes de Estados Unidos Nos 5.580.859; 5.549.127; 5.264.618; 5.703.055. La encapsulación de lípidos se realiza generalmente utilizando liposomas que son capaces de unirse de manera estable o atrapar y retener el ácido nucleico.

Los procedimientos descritos en el presente documento también se pueden usar para generar organismos modificados genéticamente no humanos. En general, en estos procedimientos, los casetes de expresión que comprenden secuencias de polinucleótidos de los diversos componentes, así como un vector de direccionamiento, se introducen en las células del cigoto para introducir específicamente en el sitio una secuencia de polinucleótidos seleccionada en una secuencia diana ADN en el genoma para generar una modificación del ADN genómico. La secuencia polinucleotídica seleccionada está presente en el vector de direccionamiento. Las modificaciones del ADN genómico suelen incluir la inserción de una secuencia de polinucleótidos, la eliminación de una secuencia de polinucleótidos o la mutación de una secuencia de polinucleótidos, por ejemplo, corrección de genes, reemplazo de genes, marcaje de genes, inserción de transgén, alteración de genes, mutación de genes, mutación de secuencias reguladoras de genes, y así sucesivamente. En un aspecto de los procedimientos para generar organismos genéticamente modificados no humanos, el organismo es un ratón. La generación de ratones transgénicos implica cinco etapas básicas (Cho A., et al., "Generation of Transgenic Mices", *Current protocols in cell biology*, (2009); CAPÍTULO. Unidad-19.11): (1) purificación de un constructo transgénico (por ejemplo, casetes de expresión que comprenden los diversos componentes de los diversos procedimientos descritos en el presente documento, así como un vector de direccionamiento); (2) recolección de cigotos donantes; (3) microinyección del constructo transgénico en el cigoto de ratón; (4) implantación de los cigotos microinyectados en ratones receptores pseudo-preñados; y (5) realización del genotipado y análisis de la modificación del ADN genómico establecido en ratones fundadores.

En otro aspecto de los procedimientos para generar organismos genéticamente modificados no humanos, el organismo es una planta. Por lo tanto, los componentes descritos en el presente documento se utilizan para efectuar la edición y manipulación génica de manera eficiente y rentable en células vegetales. En general, es preferible insertar un ADN recombinante funcional en el genoma de una planta en una ubicación no específica. Sin embargo, en ciertos casos, puede ser útil usar la integración específica del sitio para introducir un constructo de ADN recombinante en el genoma. Los vectores recombinantes para uso en plantas son conocidos en la técnica. Los vectores pueden incluir, por ejemplo, regiones de unión de andamios (SAR), orígenes de replicación y/o marcadores seleccionables.

Los procedimientos y composiciones para transformar plantas mediante la introducción de un constructo de ADN recombinante en un genoma de plantas incluyen cualquiera de una serie de procedimientos conocidos en la técnica. Un procedimiento para construir plantas transformadas es el bombardeo con microproyectiles. La transformación mediada por *Agrobacterium* es otro procedimiento para construir plantas transformadas. Alternativamente, pueden usarse otras especies no *Agrobacterium* (por ejemplo, *Rhizobium*) y otras células procariotas que pueden infectar células vegetales e introducir secuencias de nucleótidos heterólogas en el genoma de las células vegetales infectadas. Otros procedimientos de transformación incluyen la electroporación, los liposomas, la transformación con polen o virus, los productos químicos que aumentan la captación de ADN libre o el suministro de ADN libre mediante el bombardeo con microproyectiles. Los constructos de ADN de la presente invención pueden introducirse en el genoma de un huésped vegetal utilizando técnicas de transformación convencionales que son bien conocidas por los expertos en la técnica (véase, por ejemplo, "Methods to Transfer Foreign Genes to Plants," Y Narusaka, et al., cdn.intechopen.com/pdfs-wm/30876.pdf).

Si bien las realizaciones preferidas de la presente invención se han mostrado y descrito en el presente documento, será obvio para los expertos en la materia que tales realizaciones se proporcionan solo a modo de ejemplo.

15 Parte experimental

Los aspectos de la presente invención se ilustran adicionalmente en los siguientes ejemplos. Los ejemplos que no pertenecen a la invención son solo para fines ilustrativos. Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (por ejemplo, cantidades, concentraciones, cambios porcentuales, etc.) pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, la temperatura está en grados centígrados y la presión está en o cerca a la atmosférica. Debe entenderse que estos ejemplos, aunque indican algunas realizaciones de la invención, se proporcionan únicamente a modo de ilustración.

I. Uso de proteínas cas9 catalíticamente inactivas como bloqueadores de nucleasa específica del sitio

Los siguientes ejemplos 1-4 ilustran el uso de una Cas9 catalíticamente inactiva (es decir, Cas9 o dCas9 "muerta") para reducir la actividad de la nucleasa no deseada en células eucarióticas. Además, este ejemplo muestra cómo se puede identificar una secuencia espaciadora específica (para la incorporación en un ARNgu o ARNcr) que sea eficaz para bloquear la actividad de la nucleasa no deseada en células eucariotas. Cuando se usa el término ARNgu o ARN guía único, un experto en la técnica entiende que otros sistemas de polinucleótidos guía, tales como un sistema de doble guía ARNcr ARNcrtra, presentan un medio alternativo para guiar dCas9 al sitio objetivo.

Ejemplo 1

30 Producción de bloqueadores de nucleasa dCas9 y componentes de nucleasa Cas9

Componentes de ARNgu del bloqueador de nucleasa dCas9 (dCas9-NB, es decir una Cas9 que carece de actividad catalítica) complejos de ribonucleoproteína (RNP) (también llamados "complejo de ARNgu/dCas9" en este documento) y complejos de RNP de nucleasa Cas9 catalíticamente activa (también denominados "complejo de ARNgu/Cas9" en el presente documento) se produjeron por transcripción *in vitro* (por ejemplo, kit rápido de síntesis de ARN de alto rendimiento T7, New England Biolabs, Ipswich, MA) a partir de moldes de ADN de cadena doble que incorporan un promotor T7 en el extremo 5' de la secuencia de ADN. La reacción en cadena de la polimerasa (PCR) utilizando cebadores superpuestos 5' se usó para ensamblar los moldes de ADN de cadena doble para la transcripción de los componentes de ARNgu. Los componentes de ARNgu, las plantillas y los cebadores utilizados se identifican en la Tabla 1. Las secuencias de los cebadores oligonucleotídicos utilizados en el ensamblaje se presentan en la Tabla 2.

Tabla 1: Cebadores superpuestos para la generación de las plantillas de componentes de ARNgu de nucleasa dCas9-NB y Cas9

Componente	Objetivo para unión de ADN	Cebadores
ARNgu de Cas9	VEGFA	A, B, C, D, E
ARNgu de dCas9	AAVS1	A, B, C, D, F
ARNgu de dCas9	A2 alejado del objetivo del VEGFA	A, B, C, D, G
ARNgu de dCas9	A3 alejado del objetivo del VEGFA	A, B, C, D, H
ARNgu de dCas9	A4 alejado del objetivo del VEGFA	A, B, C, D, I
ARNgu de dCas9	B1 alejado del objetivo del VEGFA	A, B, C, D, J

(continuación)

Componente	Objetivo para unión de ADN	Cebadores
ARNgu de dCas9	B2 alejado del objetivo del <i>VEGFA</i>	A, B, C, D, K
ARNgu de dCas9	C1 alejado del objetivo del <i>VEGFA</i>	A, B, C, D, L
ARNgu de dCas9	C3 alejado del objetivo del <i>VEGFA</i>	A, B, C, D, M
ARNgu de dCas9	D2 alejado del objetivo del <i>VEGFA</i>	A, B, C, D, N
ARNgu de dCas9	D3 alejado del objetivo del <i>VEGFA</i>	A, B, C, D, O
*Las secuencias cebadoras de ADN se muestran en la Tabla 2		

Tabla 2: Secuencias utilizadas de cebadores de ADN

A	AAAAAAGCACCGACTCGGTGCC	SEQ ID NO: 1
B	AGTAATAATACGACTCACTATAG	SEQ ID NO: 2
C	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAAC	SEQ ID NO: 3
D	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGG ACTAGC	SEQ ID NO: 4
E	TAATACGACTCACTATAGGGTGGGGGAGTTTGCTCCGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 5
F	TAATACGACTCACTATAGGGGCCACTAGGGACAGGATGTTTTAG AGCTAGAAATAGC	SEQ ID NO: 6
G	TAATACGACTCACTATAGTGGAGGGAGTTTGCTCCTGGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 7
H	TAATACGACTCACTATAGGACGGATTTGTGGGATGGAGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 8
I	TAATACGACTCACTATAGCAGGACATTCTGACACCCCGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 9
J	TAATACGACTCACTATAGGAGGCTCCCATCACGGGGGGTTTTAG AGCTAGAAATAGC	SEQ ID NO: 10
K	TAATACGACTCACTATAGTGGGGATCACAGGTTCCCCGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 11
L	TAATACGACTCACTATAGAGAGCTCTTCTGACTACAGGTTTTAGA GCTAGAAATAGC	SEQ ID NO: 12
M	TAATACGACTCACTATAGGACCAAATGAGACCAGTCCGTTTTAG AGCTAGAAATAGC	SEQ ID NO: 13
N	TAATACGACTCACTATAGGACCAAATGAGACCAGTCCGTTTTAG AGCTAGAAATAGC	SEQ ID NO: 14
O	TAATACGACTCACTATAGCTCCTGGGGATGGAAGGGCGTTTTAG AGCTAGAAATAGC	SEQ ID NO: 15
P	CACTCTTTCCTACACGACGCTCTTCCGATCTCCAGATGGCACAT TGTCAGA	SEQ ID NO: 16

(continuación)

Q	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTAGTGACTGCCGTCTGC	SEQ ID NO: 17
R	GGAGTTCAGACGTGTGCTTCCGATCTacctggccATCTCCTTCTA	SEQ ID NO: 18
S	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCAGACCCACTGAGTCAA	SEQ ID NO: 19
T	CAAGCAGAAGACGGCATAACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTC	SEQ ID NO: 20
U	AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNACACTCTTTCCCTACACGACG	SEQ ID NO: 21

5 La reacción PCR para ensamblar el molde de ADN de ARN_g procedió de la siguiente manera: Tres cebadores de ADN "internos" (C, D, E-O, Tabla 2) estaban presentes a una concentración de 2 nM cada uno. Dos cebadores de ADN "externos" (A, B, Tabla 2) correspondientes al promotor T7 y el extremo 3' de la secuencia de ARN estaban presentes a 640 nM para dirigir la reacción de amplificación. Las reacciones PCR se realizaron utilizando el kit de PCR Kapa HiFi Hotstart^{MR} (Kapa Biosystems, Inc., Wilmington, MA) de acuerdo con las recomendaciones del fabricante. Las reacciones de ensamblaje de PCR se llevaron a cabo utilizando las siguientes condiciones de ciclos 10 térmicos: 98 °C durante 2 minutos, 35 ciclos de 15 segundos a 98 °C, 15 segundos a 62 °C, 15 segundos a 72 °C y una extensión final a 72 °C durante 2 minutos.

15 Se transcribieron aproximadamente 0,25-0,5 µg del molde de ADN para los componentes de ARN_g utilizando el kit de síntesis de ARN de alto rendimiento T7 (New England Biolabs, Ipswich, MA) durante aproximadamente 16 horas a 37 °C. Las reacciones de transcripción fueron tratadas con ADNasa I (New England Biolabs, Ipswich, MA). La calidad del ARN transcrito se verificó mediante electroforesis capilar en un analizador de fragmentos (Advanced Analytical Technologies, Inc., Ames, IA). Las secuencias del componente ARN_g de Cas9 y dCas9-NB fueron las siguientes:

Tabla 3: Secuencias del componente ARN_g de Cas9 y dCas9-NB

ADN diana	Secuencia de ARN (5' a 3')
<i>VEGFA</i>	GGGUGGGGGGAGUUUGCUCCGUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGG UGCUUUUUUU (SEQ ID NO:22)
<i>AAVS1</i>	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:23)
A2 alejado del objetivo del <i>VEGFA</i>	GUGGAGGGAGUUUGCUCCGUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGG UGCUUUUUUU (SEQ ID NO:24)
A3 alejado del objetivo del <i>VEGFA</i>	GGACGGAUUUGUGGAUGGAGUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGG UGCUUUUUUU (SEQ ID NO:25)
A4 alejado del objetivo del <i>VEGFA</i>	GCAGGACAUUCUGACACCCCGUUUUAGAGCUAGAAAUAGCAAGUUA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:26)
B1 alejado del objetivo del <i>VEGFA</i>	GGAGGCUCCCAUCACGGGGGUUUUUAGAGCUAGAAAUAGCAAGUUA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:27)

(continuación)

ADN diana	Secuencia de ARN (5' a 3')
B2 alejado del objetivo del <i>VEGFA</i>	GUGGGGAUCACAGGUUCCCCGUUUUAGAGCUAGAAAUAGCAAGUAAA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:28)
C1 alejado del objetivo del <i>VEGFA</i>	GAGAGCUCUUCUGACUACAGGUUUUAGAGCUAGAAAUAGCAAGUAAA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:29)
C3 alejado del objetivo del <i>VEGFA</i>	GGACCAAUGAGACCAGUCCGUUUUAGAGCUAGAAAUAGCAAGUAAA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUUU (SEQ ID NO:30)
D2 alejado del objetivo del <i>VEGFA</i>	GCCCAUUAUGAUAGGGAGGGGUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGG UGC UUUUUUUU (SEQ ID NO:31)
D3 alejado del objetivo del <i>VEGFA</i>	GCUCCUGGGGAUUGGAAGGGCGUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGG UGC UUUUUUUU (SEQ ID NO:32)

Los componentes proteicos de los RNP Cas9 y dCas9-NB se expresaron a partir de vectores de expresión bacterianos en *E. coli* (BL21 (DE3)) y se purificaron utilizando cromatografía de afinidad, intercambio iónico y exclusión por tamaño de acuerdo con los procedimientos descritos en Jinek et al., 2012. La secuencia codificante para Cas9 de *S. pyogenes* incluyó las dos secuencias de localización nuclear (NLS) en el extremo terminal C. La variante dCas9 de Cas9 etiquetada con NLS, en la que se mutaron los residuos del sitio activo de ambos dominios de nucleasa (Jinek, et al., 2012), se prepararon mediante la introducción de mutaciones en la secuencia codificante de Cas9 de *S. pyogenes* por mutagénesis dirigida al sitio (kit de mutagénesis dirigida al sitio Q5, New England Biolabs, Ipswich, MA). Este procedimiento para la producción de los RNP Cas9 y/o dCas9-NB se puede aplicar a la producción de otras RNP Cas9 y/o dCas9-NB como se describe en este documento.

Ejemplo 2

Análisis de secuenciación profunda para la detección de modificaciones objetivo en células eucariotas

Este ejemplo ilustra el uso de un secuenciador MiSeq (Illumina, San Diego, CA) para el análisis de secuenciación profunda para evaluar y comparar la escisión del ADN (como se infiere de la unión de extremos no homólogos, o NHEJ) de secuencias no deseadas de nucleasa Cas9 seleccionada en presencia y ausencia de dCas9-NB. En este ejemplo, Cas9 fue dirigido por un ARN_g específico a una secuencia (GGGTGGGGGAGTTTGCTCCTGG, SEQ ID NO: 82) dentro del factor de crecimiento endotelial vascular A (*VEGFA*) de un gen humano. dCas9 se dirigió hacia una secuencia no deseada (GGATGGAGGGAGTTTGCTCCTGG, SEQ ID NO: 83) que se sabe que es el objetivo de la RNP nucleasa Cas9 para prevenir la escisión no deseada así como una secuencia (GGGGCCCCTAGGGACAGGATTGG, SEQ ID NO: 84) dentro del locus de control, sitio de integración de virus adenoasociado 1 (AAVS1).

A. Transfección de las RNP Cas9/dCas9-NB:

Para ensamblar las RNP Cas9 y dCas9, se incubaron 1,3 µl de ARN_g (que corresponde a aproximadamente 1-9 µg o aproximadamente 25-250 pmol) durante 2 minutos a 95 °C y luego se dejó equilibrar a temperatura ambiente durante aproximadamente 5 minutos. Posteriormente, Cas9 y dCas9 se mezclaron con un ARN_g correspondiente para formar las RNP en un tampón de reacción (HEPES 20 mM, pH 7,5, KCl 100 mM, MgCl₂ 5 mM, glicerol al 5%). Se combinaron 20 pmoles de Cas9 con el ARN_g objetivo, y se combinaron 0 o 20 pmoles de dCas9 con los ARN_g alejados del objetivo, y se ensamblaron las RNP funcionales mediante incubación a 37 °C durante 10 min. Finalmente, se combinaron 20 pmoles de RNP Cas9 con 0 (es decir, solo el componente de ARN_g de dCas9-NB) o 20 pmoles de RNP dCas9 inmediatamente antes de la transfección a las células. Los experimentos se realizaron por triplicado.

Los complejos de RNP Cas9/dCas9-NB se transfectaron en células K562 (ATCC, Manassas, VA), utilizando el sistema Nucleofector^{MR} de 96 pocillos Shuttle (Lonza, Allendale, NJ) y el siguiente protocolo: los complejos de RNP se dispensaron en un volumen final de 5 µl en pocillos individuales de una placa de 96 pocillos. Las células K562 suspendidas en medio se transfirieron del matraz de cultivo a un tubo cónico de 50 ml, luego las células se sedimentaron por centrifugación durante 3 minutos a 200 x g, el medio de cultivo se aspiró y se lavó una vez con PBS libre de calcio y magnesio. Las células K562 se sedimentaron luego por centrifugación durante 3 minutos a 200

x g, se aspiró el PBS y el sedimento celular se resuspendió en 10 ml de PBS libre de calcio y magnesio.

- Se contaron las células K562 usando el contador de células automatizado Countess® II (Life Technologies, Grand Island, NY). Se transfirieron $4,2 \times 10^7$ células a un tubo de 50 ml y se sedimentaron. La PBS se aspiró y las células se resuspendieron en una solución de Nucleofector^{MR} SF (Lonza, Allendale, NJ) a una densidad de 1×10^7 células/ml. Luego se agregaron 20 μ l de la suspensión celular a los pocillos individuales que contenían 5 μ l de complejos de RNP y todo el volumen se transfirió a los pocillos de una placa de Nucleocuvette^{MR} de 96 pocillos (Lonza, Allendale, NJ). La placa se cargó en el Nucleofector^{MR} de 96 pocillos Shuttle^{MR} (Lonza, Allendale, NJ) y las células se nucleofectaron utilizando el programa 96-FF-120 Nucleofector^{MR} (Lonza, Allendale, NJ). Después de la nucleofección, se agregaron 80 μ l del medio de Dulbecco modificado de Iscove (IMDM, Life Technologies, Grand Island, NY), complementado con un FBS al 10% (Fisher Scientific, Pittsburgh, PA) y complementado con penicilina y estreptomina (Life Technologies, Grand Island, NY), a cada pocillo y se transfirieron 50 μ l de la suspensión celular a una placa de cultivo celular de 96 pocillos que contenía 150 μ l de medio de cultivo completo IMDM precalentado. La placa se transfirió luego a una incubadora de cultivo de tejidos y se mantuvo a 37 °C en CO₂ al 5% durante aproximadamente 48 horas.
- Se aisló ADN genómico (ADNg) de células K562 48 horas después de la transfección de Cas9/dCas9-NB utilizando 50 μ l de solución de extracción de ADN QuickExtract (Epicenter, Madison, WI) por pocillo, seguido de incubación a 37 °C durante 10 minutos. Se agregaron 50 μ l de agua a las muestras, y luego se incubaron a 75 °C durante 10 minutos y 95 °C durante 5 minutos para detener la reacción. Se almacenó el ADNgu se almacenó a -20 °C hasta su posterior procesamiento.

20 B. Preparación de la biblioteca de secuenciación:

Utilizando el ADNgu previamente aislado, se realizó una primera PCR usando Q5 Hot Start High-Fidelity 2X Master Mix^{MR} (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 μ M cada uno, 3,75 μ l de ADNgu en un volumen final de 10 μ l y se amplificó a 98 °C durante 1 minuto, 35 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 min. La reacción PCR se diluyó 1:100 en agua. Los cebadores específicos del objetivo se muestran en la Tabla 4:

Tabla 4: Cebadores específicos del objetivo utilizados para la secuenciación

Objetivo	Cebadores
VEGFA objetivo	P,Q
VEGFA 1 alejado del objetivo	R,S
*Las secuencias del cebador de ADN se muestran en la Tabla 2	

Se configuró una segunda PCR de 'código de barras' utilizando cebadores únicos para cada muestra que facilitan la secuenciación múltiple (oligonucleótidos T y U en la Tabla 2, en la que una única secuencia índice de 8 pb, indicada por "NNNNNNNN (SEQ ID NO: 33)" permitió la desmultiplexación de cada amplicón durante el análisis de secuencia).

La segunda PCR se realizó utilizando Q5 Hot Master High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 μ M cada uno, 1 μ l de una PCR diluida primero 1:100, en un volumen final de 10 μ l y se amplificó a 98 °C durante 1 minuto, 12 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 min. Las reacciones de PCR se agruparon en un solo tubo de microcentrifuga para la limpieza de amplicones para secuenciación basada en perlas SPRIselect (Beckman Coulter, Pasadena, CA).

A los amplicones reunidos, se les agregaron 0,9 volúmenes de perlas SPRIselect, se mezclaron y se incubaron a temperatura ambiente (RT) durante 10 minutos. El tubo de microcentrifuga se colocó en un soporte de tubos magnéticos (Beckman Coulter, Pasadena, CA) hasta que la solución se aclaró. El sobrenadante se eliminó y se desechó, y las perlas residuales se lavaron con 1 volumen de etanol al 85% y se incubaron a temperatura ambiente durante 30 s. Después de la incubación, se aspiró el etanol y las perlas se secaron al aire a temperatura ambiente durante 10 min. Luego, el tubo de microcentrifuga se retiró del soporte magnético y se agregaron 0,25x volúmenes de agua (Qiagen, Venlo, Limburgo) a las perlas, se mezclaron vigorosamente y se incubaron durante 2 min a RT. Luego se centrifugó el tubo de microcentrifuga en una microcentrifuga para recoger el contenido del tubo, y luego se devolvió al imán, se incubó hasta que la solución se aclaró y el sobrenadante que contenía los amplicones purificados se dispensó en un tubo limpio de microcentrifuga. La biblioteca de amplicones purificada se cuantificó utilizando el sistema Nanodrop^{MR} 2000 (Thermo Fisher Scientific, Wilmington, DE).

La biblioteca de amplicones se normalizó a una concentración de 4 nM calculada a partir de la absorbancia óptica a 260 nm (Nanodrop^{MR}, Thermo Fisher Scientific, Wilmington, DE) y el tamaño de los amplicones. La biblioteca se analizó en un secuenciador MiSeq con el kit de reactivos de MiSeq v2, 300 ciclos (Illumina, San Diego, CA), con dos

procesos de finalización pareados de 151 ciclos más dos lecturas de índice de ocho ciclos.

C. Análisis de datos de secuenciación profunda:

La identidad de los productos en los datos de secuenciación se analiza con base en la indexación de secuencias de código de barras adaptadas a los amplicones en la segunda ronda de PCR. Se utilizó un guión computacional para procesar los datos de MiSeq ejecutando las siguientes tareas:

1. Las lecturas se alinearon con el genoma humano (compilación GRCh38/38) utilizando el software Bowtie (bowtie-bio.sourceforge.net/index.shtml).
 2. Las lecturas alineadas se compararon con los loci de tipo silvestre; las lecturas que no se alineaban con ninguna parte de los loci fueron descartadas.
 3. Las lecturas que coinciden con la secuencia de tipo silvestre se registraron. Las lecturas con inserciones o supresiones (alrededor de 5 pb del sitio de corte de Cas9) se clasificaron por el tipo de inserciones o supresiones y se contabilizaron.
 4. Las lecturas de inserciones o supresiones totales se dividieron por la suma de las lecturas de tipo silvestre y las lecturas de inserciones o supresiones para obtener lecturas del porcentaje mutado.
- La Figura 8 muestra los efectos del Cas9-NB en la edición deseada de ARN_{gu}/Cas9 del *VEGFA* en el locus del *VEGFA*. Como se puede ver a partir de los datos de la figura, la adición de un dCas9-NB dirigido al locus objetivo del *VEGFA* inhibe la edición deseada, mientras que los dCas9-NB dirigidos a distintas regiones no tienen un efecto significativo (n = 3, las barras de error muestran la desviación estándar, *p < 0,05, prueba t de Student (dos colas) comparando 3 con 4).
- La Figura 9 muestra los efectos de dCas9-NB en la edición no deseada de ARN_{gu}/Cas9 del *VEGFA* en un locus conocido alejado del objetivo del *VEGFA* en el cromosoma humano 15. Como se puede observar a partir de los datos en la figura, la adición de un dCas9-NB al locus alejado del objetivo, ya sea mediante el ARN_{gu} objetivo del *VEGFA* o un ARN_{gu} dirigido específicamente al locus alejado del objetivo del cromosoma 15, perjudica la escisión, mientras que el dCas9-NB dirigido a distintas regiones no tiene un efecto significativo (n = 3, las barras de error muestran la desviación estándar, *p < 0,05, prueba t de Student (dos colas) comparando 3 con 4, 5 con 6, 7 con 8).

En la Tabla 5 se muestra una descripción de las muestras utilizadas en estos experimentos y las Figuras 8 y 9:

Tabla 5: Descripciones de muestras para las Figuras 8 y 9

Muestra	Descripción
1	ARN _{gu} del <i>AAVS1</i> , 0 pmol de dCas9
2	ARN _{gu} del <i>AAVS1</i> , 20 pmol de dCas9
3	ARN _{gu} del <i>VEGFA</i> , 0 pmol de dCas9
4	ARN _{gu} del <i>VEGFA</i> , 20 pmol de dCas9
5	ARN _{gu} A2 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
6	ARN _{gu} A2 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
7	ARN _{gu} A3 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
8	ARN _{gu} A3 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
9	ARN _{gu} A4 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
10	ARN _{gu} A4 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
11	ARN _{gu} B1 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
12	ARN _{gu} B1 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
13	ARN _{gu} B2 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
14	ARN _{gu} B2 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9

(continuación)

Muestra	Descripción
15	ARNgu C1 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
16	ARNgu C1 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
17	ARNgu C3 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
18	ARNgu C3 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
19	ARNgu D2 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
20	ARNgu D2 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9
21	ARNgu D3 alejado del objetivo del <i>VEGFA</i> , 0 pmol de dCas9
22	ARNgu D3 alejado del objetivo del <i>VEGFA</i> , 20 pmol de dCas9

5 Siguiendo la guía de la presente memoria descriptiva y los ejemplos, el experto en la técnica puede practicar el análisis de secuenciación profunda descrito en este ejemplo con otros complejos de RNP Cas9/dCas9 (es decir, ensamblados con ARNgu distintos y/o relaciones distintas de Cas9, dCas9 y ARNgu).

Ejemplo 3

Identificación de loci alejados del objetivo de RNP Cas9

10 Este ejemplo ilustra el procedimiento a través del cual se pueden identificar los sitios de nucleasa Cas9 alejados del objetivo. El procedimiento presentado en el presente documento es adaptado de Tsai et. al., "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPRCas nucleases" Nat Biotechnol., febrero de 2015; 33 (2): 187-97.

A. Identificación de un sitio objetivo de interés:

15 Un locus dado en un genoma de interés (es decir, un genoma humano) se analiza utilizando enfoques bioinformáticos conocidos por los expertos en la técnica para identificar sitios objetivo de Cas9. Se selecciona un sitio objetivo de 20 pares de bases, seguido de un motivo adyacente protoespaciador NGG (PAM), para direccionamiento de la nucleasa.

B. Ensamble los componentes de GUIDE-Seq:

20 Se obtienen oligos (Integrated DNA Technologies, Coralville, IA) para generar un oligodesoxinucleótido como de cadena doble (DONcd) que se utilizará para el procedimiento GUIDE-Seq. El DONcd contiene enlaces de fosfotioato en los extremos 5' de ambas cadenas de ADN. El DONcd se ensambla incubando los dos oligos en un tampón de hibridación (es decir, Tris 10 mM, pH 7,5-8,0, NaCl 50 mM, EDTA 1 mM) durante 3 minutos a 95 °C y dejando que los oligos se enfríen hasta temperatura ambiente.

C. Transfección de componentes de GUIDE-Seq:

25 Las células de una especie de interés (es decir, células humanas) se obtienen de un repositorio comercial (es decir, ATCC, DSMZ). Las células se cultivan hasta una densidad apropiada para la transfección. Las células se transfectan con un complejo de ARNgu/proteína Cas9 y el oligo donante de ADN a través de procedimientos conocidos por los expertos en la técnica (es decir, nucleofección o transfección de lípidos del plásmido de ADN que codifica los componentes de RNP, así como DONcd).

D. Preparación y análisis de la biblioteca de secuenciación:

30 El ADNg se recolecta 48 horas después de la transfección celular y se purifica utilizando Agencourt DNAdvance (Beckman Coulter, Pasadena, CA). El ADNg purificado se fragmenta con procedimientos conocidos por los expertos en la técnica (es decir, corte mecánico mediante sonicación o corte enzimático con NEBfragmentase, (New England Biolabs, Ipswich, MA)) a una longitud promedio de 500 pares de bases, luego se repara el extremo, se añade una cola de poliA y se liga a adaptadores. La PCR con cebadores complementarios a la etiqueta de DONcd y las secuencias adaptadoras de secuenciación de illumina (Illumina, San Diego, CA), respectivamente, se utilizan para el enriquecimiento con el objetivo. La biblioteca enriquecida con el objetivo se secuencia usando el secuenciador MiSeq (Illumina, San Diego, CA). Las lecturas se asignan nuevamente al genoma de la especie respectiva y la

cobertura de lectura se calcula utilizando BedTools (bedtools.readthedocs.org/en/latest/). Integrative Genomics Viewer (IGV, broadinstitute.org/igv/) se usa para mapear la posición inicial (5') y final (3') de las lecturas, y los picos se llaman usando MACS2 (pypi.python.org/pypi/MACS2). Los datos de secuenciación se utilizan para confirmar que un locus genómico putativo es una secuencia no deseada candidata. Siguiendo los lineamientos de los presentes ejemplos, un experto en la técnica puede realizar la identificación de nuevos loci alejados del objetivo.

Ejemplo 4

Bloqueo de dCas9 no deseada con ARN guía únicos truncados (ARNgtru)

Este ejemplo ilustra procedimientos en los que los dCas9-NB se pueden ensamblar con guías truncadas. El procedimiento presentado aquí es adaptado de Fu Y et. al., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs", *Nat Biotechnol.* Marzo de 2014; 32 (3): 279-84. Se ha demostrado que los ARN guía únicos (ARNgtru) de 17-18 nt tienen una mayor especificidad en relación con los ARNgu de 20 nt. Por lo tanto, un dCas9-NB ensamblado con un ARNgtru puede ser dirigido directamente a un motivo genómico y PAM de un locus alejado del objetivo para reducir la edición no deseada mientras que tiene una inhibición mínima de la edición deseada.

A. Diseño de ARNgtru para permitir el bloqueo de nucleasa alejado del objetivo mediado por dCas9:

Usando los procedimientos descritos en el Ejemplo 3 de este documento, se identifica un locus genómico alejado del objetivo dado (es decir, una secuencia espaciadora). A continuación, se diseña un ARNgtru para dirigirse a dicha ubicación no deseada en el genoma. La RNP ARNgtru/dCas9 puede apuntar a una secuencia contenida completamente dentro del motivo alejado del objetivo, o puede apuntar a una secuencia que se superpone parcialmente con el motivo alejado del objetivo.

B. Producción de componentes bloqueadores de nucleasas dCas9:

dCas9 se ensambla con un ARNgtru corto (es decir, 17 nt) y Cas9 se ensambla con un ARNgu (es decir, 20 nt) para producir RNP funcionales. Los componentes de ARN se transcriben a partir de moldes de ADN que incorporan un promotor T7 en el extremo 5' como se describe en la sección experimental de este documento. Las proteínas dCas9 (D10A, H840A) y Cas9 se expresan de forma recombinante en RNP de *E. coli* se ensamblan mediante la incubación de componentes de proteína y ARN juntos a 37 °C durante 10 minutos.

C. Transfección de ARNgtru que contiene dCas9-NB y ARNgu que contiene RNP Cas9:

Las células de especies de interés se obtienen de un repositorio comercial (es decir, ATCC, DSMZ). Las células se cultivan hasta un nivel de confluencia que permite la transfección. ARNgtru complejados con dCas9 se mezclan con ARNgu ensamblados con Cas9. A continuación, la mezcla se transfecta en una línea celular de interés utilizando procedimientos conocidos por los expertos en la técnica (es decir, nucleofección o transfección de lípidos) como se describe en el Ejemplo 1 de este documento.

D. Preparación de la biblioteca de secuenciación:

ADNg se recoge luego 48 horas más tarde usando Quick Extract (Epicenter, Madison, WI) según las instrucciones del fabricante. Se utilizan dos rondas de PCR, como se describe en el Ejemplo 1 de este documento, para amplificar y codificar mediante código de barras la región genómica seleccionada por el ARNgtru/dCas9-NB. Los oligos y dímeros adaptadores se eliminan mediante la limpieza de la biblioteca de secuenciación basada en perlas SPRIselect (Beckman Coulter, Pasadena, CA). La concentración de la biblioteca de secuenciación está determinada por el sistema Nanodrop^{MR} 2000 (Thermo Scientific, Wilmington DE).

E. Análisis de secuenciación profunda:

La biblioteca se analiza en el secuenciador MiSeq de la siguiente manera:

1. Las lecturas se alinean con el genoma humano (GRCh38/38 compilado) utilizando el software Bowtie (bio.sourceforge.net/index.shtml).
2. Las lecturas alineadas se comparan con los loci de tipo silvestre; las lecturas que no se alinean con ninguna parte de los loci se descartan.
3. Las lecturas que coinciden con la secuencia de tipo silvestre se contabilizan. Las lecturas con inserciones o supresiones se clasifican por el tipo de inserciones o supresiones y se contabilizan.
4. Las lecturas de inserciones o supresiones totales se dividen por la suma de las lecturas de tipo silvestre y las lecturas de inserciones o supresiones son lecturas con un porcentaje de mutación.

II. Posicionamiento directo por dCas9 de donantes de reparación dirigidos por homología

50

Ejemplo 5

Uso de dCas9 para colocar los nucleótidos de los donantes de homología cerca de un sitio específico para aumentar la eficiencia de la reparación dirigida por homología (HDR).

Este sistema consiste en una endonucleasa específica del sitio (por ejemplo, Cas9 complejada con un ARN_g) que se dirige a una secuencia diana ADN de interés, y una o más moléculas de dCas9 catalíticamente inactivas complejadas con un ARN_g que se dirige a secuencias de DNA adyacentes al sitio de corte (véanse las Figs. 7A y 7B). Estas moléculas de dCas9 también están unidas a una molécula de HDR (es decir, ADN_{cd}, ADN_{cs}, ARN, un plásmido, o similares). La dCas9 unida se utiliza para colocar la molécula donante en una orientación que aumente la probabilidad de que la molécula donante se incorpore en el sitio objetivo a través de HDR, introduciendo así un cambio deseado en la secuencia objetivo.

10 A. Constructos de ADN y ARN:

Los oligonucleótidos se piden a los fabricantes (por ejemplo, Integrated DNA Technologies, Coralville, IA; o Eurofins, Luxemburgo). Los constructos de transcripción de ARN_g se ensamblan mediante la reacción en cadena de la polimerasa (PCR).

15 Los cebadores para constructos de transcripción de ARN_g consisten en un cebador que contiene una secuencia promotora T7 5', un cebador que contiene una secuencia espaciadora única, cebadores que contienen la cadena principal TRCR de ARN_g y un cebador inverso que puede contener una secuencia complementaria al donante de homología para unir el donante al extremo 3' del ARN_g.

20 Los constructos de transcripción de ARN_g T7 se amplifican por PCR. Dos cebadores externos (el oligo directo contiene el oligo promotor T7; los oligos inversos contienen el extremo 3' de la cadena principal del ARN_g o la secuencia complementaria del donante de homología para la unión) están presentes en la reacción de PCR a 640 nM. El espaciador único y los oligos de la cadena principal de ARN_g están presentes a 2 nM. Las reacciones PCR se realizan utilizando Q5 Hot Start High-Fidelity 2X Master Mix^{MR} (New England Biolabs, Ipswich, MA) siguiendo las recomendaciones del fabricante. La PCR del montaje del constructo de transcripción de ARN_g T7 por PCR se lleva a cabo utilizando las siguientes condiciones de ciclos térmicos: 98 °C durante 2 minutos, 29 ciclos de 98 °C durante 25 segundos, 62 °C durante 20 segundos y 72 °C durante 15 segundos, seguido por una extensión final de 72 °C durante 2 minutos. Los constructos de ADN se evalúan mediante electroforesis capilar (análizador de fragmentos, Advanced Analytical Technologies, Ames, IA).

30 Los componentes de ARN se producen a través de transcripción *in vitro* (kit de síntesis de ARN de alto rendimiento T7 Quick, New England Biolabs, Ipswich, MA) a partir de un molde de ADN de cadena doble. El ARN se trata luego con DNasa I (New England Biolabs, Ipswich, MA) para eliminar cualquier ADN de cadena doble y se incuba a 37 °C durante 10 minutos. Luego se agrega EDTA 0,5 M a las reacciones de transcripción y se incuba a 75 °C durante 10 minutos para inactivar la DNasa I.

35 Los donantes de homología se ordenan como oligos de ADN de cadena sencilla de aproximadamente 90 nucleótidos de longitud. Los donantes de homología son complementarios a la secuencia de codificación y se diseñan para centrarse en el sitio de corte con el PAM reemplazado con un sitio de la enzima de restricción *EcoR*I y los brazos de homología de aproximadamente 42 nucleótidos de longitud que coinciden con la secuencia objetivo.

B. Generación del complejo de ARN_g/Cas9 y ARN_g/dCas9:

40 Cas9 catalíticamente activa y dCas9 catalíticamente inactiva de *S. pyogenes* se marcan en el extremo terminal C con dos secuencias de localización nuclear (NLS) y se expresan de forma recombinante en *E. coli*. Todos los ARN_g y ARN_g unidos se incuban durante 2 minutos a 95 °C, se retiran del termociclador y se dejan equilibrar a temperatura ambiente. Los complejos de ribonucleoproteína (RNP) Cas9 (también denominados "complejo de ARN_g/Cas9" y "complejo de ARN_g/dCas9" en este documento) se configuran por triplicado con Cas9 2 μM o dCas9 2 μM, ARN_g 6 μM o ARN_g unido 6 μM y oligo donante 2 μM en el tampón de unión (HEPES 20 mM, KCl 100 mM, MgCl₂ 5 mM, DTT 1 mM y glicerol al 5% a pH 7,4) en un volumen de 6 μl. A continuación, se deja que las RNP se unan a 37 °C durante 10 minutos. Después de la hibridación, las uniones del donante de RNP Cas9 y RNP dCas9 se pueden combinar hasta un volumen final de 12 μl.

C. Condiciones experimentales:

Las condiciones experimentales para el uso de diversos aspectos de la invención se describen a continuación y se ilustran en la Figura 10.

- 50
- 1) Sin control de transfección - las células no son electroporadas.
 - 2) Cas9 + ARN_g estándar - las células se transfectan con ARN_g y Cas9 catalíticamente activa que se dirige al locus de interés.
 - 3) Cas9 + ARN_g estándar para el sitio adyacente al locus objetivo - las células se transfectan con ARN_g y Cas9 catalíticamente activo que se dirige al sitio adyacente del locus para determinar si el espaciador y el PAM

en ese sitio proporcionan una buena actividad de escisión en ese sitio

4) Control único del donante - las células se transfectan con el polinucleótido donante para determinar si hay incorporación a través de HDR en el sitio

5) ARN_g/dCas9 unido y transfección de donante - las células se transfectan con el polinucleótido donante y con complejos de ARN_g/dCas9 unidos para el sitio adyacente al locus objetivo.

6) ARN_g/Cas9 unido en el locus objetivo - las células se transfectan con un complejo de ARN_g/Cas9 unido catalíticamente activo con un polinucleótido donante al locus objetivo.

7) ARN_g/Cas9 estándar y transfección del donante - las células se transfectan con el complejo de ARN_g/Cas9 catalíticamente activo y el polinucleótido donante. Esto determinará las tasas de incorporación de HDR para un experimento de HDR estándar.

8) ARN_g/Cas9 estándar y ARN_g/dCas9 unido (en el sitio) y transfección del donante - las células se transfectan con el complejo de ARN_g/Cas9 catalíticamente activo que se dirige al locus objetivo y con ARN_g/dCas9 unido que se dirige al locus adyacente objetivo y al polinucleótido donante.

9) ARN_g/Cas9 estándar y ARN_g/dCas9 unido (dos sitios) y transfección del donante - las células se transfectan con un complejo de ARN_g/Cas9 catalíticamente activo dirigido al locus objetivo y con ARN_g/dCas9 unido que se dirige hacia arriba y hacia abajo del locus adyacente objetivo y el polinucleótido donante. Los dos se unen en los dos complejos de ARN_g/dCas9 que estiran el polinucleótido donante a través del rompimiento de cadena doble y vuelven esa región disponible para HDR.

D. Cultivo celular y transfecciones:

Las células K562 (ATCC, Manassas, VA) se cultivan en suspensión en medio IMDM complementado con FBS al 10% y penicilina al 1% y estreptomycin a 37°C con humedad del 100%. Las células K562 se transfectan utilizando el sistema Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ). Las RNP y las RNP dCas9 se disponen en una placa de 96 pocillos con 2 µl de RNP Cas9 o 4 µl de Cas9 y RNP dCas9 combinados. Las células K562 se transfieren a un tubo de centrifuga cónico de 50 ml y se centrifugan a 200 x g durante 3 minutos. El medio se aspira y el sedimento celular se lava en PBS libre de calcio y magnesio. Las células se centrifugan una vez más y luego se resuspenden en tampón Nucleofector SF (Lonza, Allendale, NJ) a una concentración de 1x10⁷ células/ml. Se agregan 20 µl de esta suspensión celular a la RNP en la placa de 96 pocillos, se mezcla, y luego todo el volumen se transfiere a una placa de Nucleocuvette^{MR} de 96 pocillos (Lonza, Allendale, NJ). Luego, la placa se carga en el sistema Nucleofector® 96-well Shuttle^{MR} (Lonza, Allendale, NJ) y las células se nucleofectan utilizando el programa 96-FF-120 Nucleofector^{MR} (Lonza, Allendale, NJ). Inmediatamente después de la nucleofección, se añaden 80 µl de medio completo IMDM a cada pocillo de la placa Nucleocuvette^{MR} de 96 pocillos. El contenido completo del pozo se transfiere luego a una placa de cultivo de tejidos de 96 pocillos que contiene 100 µl de medio IMDM completo. Las células se cultivan a 37 °C con condiciones de humedad del 100% durante 48 horas.

Después de 48 horas, las células K562 se centrifugan a 500 x g durante 5 minutos y se elimina el medio. Las células se lavan 1 vez en PBS libre de calcio y magnesio. Los sedimentos celulares se resuspenden luego en 50 µl de soluciones de extracción de ADN QuickExtract (Epicenter, Madison, WI). Las muestras de ADN_g obtenidas se incuban a 37 °C durante 10 minutos, 65 °C durante 6 minutos y 95 °C durante 3 minutos para detener la reacción. Las muestras de ADN_g se diluyen luego con 50 µl de agua y se almacenan a -20 °C.

Este ADN_g se amplifica por PCR utilizando Q5 Hot Start High-Fidelity 2X Master Mix^{MR} (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 µM cada uno, 3,75 µl de ADN_g en un volumen final de 10 L y amplificado a 98 °C durante 1 minuto, 35 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 min. La reacción PCR se diluyó 1:100 en agua.

Una segunda PCR de "código de barras" se configura utilizando cebadores únicos para cada muestra, lo que facilita la secuenciación múltiplex. La segunda PCR se realiza utilizando Q5 Hot Master High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 µM cada uno, 1 µl de una PCR diluida primero 1:100, en un volumen final de 10 µl y se amplificó a 98 °C durante 1 minuto, 12 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 minutos.

E. Limpieza con SPRIselect:

Las reacciones de PCR se agrupan en un solo tubo de microcentrifuga para la limpieza de amplicones para secuenciación basada en perlas SPRIselect^{MR} (Beckman Coulter, Pasadena, CA).

A los amplicones agrupados, se les agregan 0,9 volúmenes de perlas SPRIselect^{MR}, y se mezclan y se incuban a temperatura ambiente (RT) durante 10 minutos. El tubo de microcentrifuga se coloca en un soporte de tubo magnético (Beckman Coulter, Pasadena, CA) hasta que la solución se haya aclarado. El sobrenadante se elimina y se desecha, y las perlas residuales se lavan con 1 volumen de etanol al 85% y se incuban a temperatura ambiente

durante 30 s. Después de la incubación, se aspira el etanol y las perlas se secan al aire a temperatura ambiente durante 10 min. El tubo de microcentrífuga se retira del soporte magnético y se agregan 0,25x volúmenes de tampón Qiagen EB (Qiagen, Venlo, Países Bajos) a las perlas, se mezcla vigorosamente y se incuba durante 2 minutos a temperatura ambiente. El tubo de microcentrífuga se devuelve al imán, se incuba hasta que la solución se ha aclarado y el sobrenadante que contiene los amplicones purificados se dispensa en un tubo limpio de microcentrífuga. La biblioteca de amplicones purificados se cuantifica utilizando el sistema Nanodrop^{MR} 2000 (Thermo Fisher Scientific, Wilmington DE) y se analiza la calidad de la biblioteca utilizando el sistema Fragment Analyzer^{MR} (Advanced Analytical Technologies, Inc., Ames, IA) y el DNF-910 dsDNA Reagent Kit^{MR} (Advanced Analytical Technologies, Inc. Ames, IA).

10 **F. Configuración de secuenciación profunda:**

La biblioteca de amplicones se normaliza hasta una concentración 4 nM como la calculada a partir de los valores de Nanodrop y el tamaño de los amplicones. La biblioteca se analiza en un secuenciador MiSeq con el MiSeq Reagent Kit v2^{MR}, 300 ciclos (Illumina, San Diego, CA), con dos procesos de finalización pareados de 151 ciclos más dos lecturas de índice de ocho ciclos.

15 **G. Análisis de datos de secuenciación profunda:**

La identidad de los productos en los datos de secuenciación se analiza con base en la indexación de secuencias de código de barras adaptadas a los amplicones en la segunda ronda de PCR. Se utilizó un guión computacional para procesar los datos de MiSeq ejecutando las siguientes tareas:

- 20 1. Las lecturas se alinearon con el genoma humano (compilación GRCh38/38) utilizando el software Bowtie (bowtie-bio.sourceforge.net/index.shtml).
2. Las lecturas alineadas se compararon con los loci de tipo silvestre; las lecturas que no se alineaban con ninguna parte de los loci fueron descartadas.
3. Las lecturas que coinciden con la secuencia de tipo silvestre se registraron. Las lecturas con inserciones o supresiones se clasificaron por el tipo de inserciones o supresiones y se contabilizaron.
- 25 4. Las lecturas de inserciones o supresiones totales se dividieron por la suma de las lecturas de tipo silvestre y las lecturas de inserciones o supresiones para obtener lecturas del porcentaje mutado.

Estos datos se analizan luego para determinar si los polinucleótidos donantes unidos a ARN_g/dCas9 aumentan la eficacia de HDR en comparación con los polinucleótidos donantes difundidos en forma pasiva.

30 **III. Uso de mutantes de nickasa Cas9 para aumentar la eficiencia de la reparación dirigida por homología como una fracción de eventos totales de reparación**

Ejemplo 6

Uso de nickasas Cas9 en tándem para dirigir la reparación dirigida por homología en los sitios de escisión en células eucarióticas

35 Este ejemplo ilustra el uso de un mutante de nickasa Cas9 en el que un dominio de nucleasa está inactivado (Cas9D10A) para emplear preferentemente vías de reparación dirigidas por homología (HDR) y bloquear vías de reparación mutagénica en sitios de rompimiento en células eucariotas. En este ejemplo, Cas9D10A se usa con dos ARN guía únicos específicos (ARN_g) que suministran la nickasa a dos sitios en la misma cadena separados por 30-60 nucleótidos. Las secuencias espaciadoras se eligieron de las secuencias disponibles en el ADN genómico humano, de modo que cada uno de los dos ARN_g apuntaría a Cas9 en una ubicación en cualquier lado de la región deseada para su modificación.

Producción de nickasa Cas9D10A y componentes de nucleasa Cas9:

45 Los componentes de ARN_g de los complejos de ribonucleoproteína Cas9D10A (RNP) (también denominados "complejos de ARN_g/nickasa Cas9" en este documento) y los complejos de RNP de nucleasa Cas9 catalíticamente activos (también denominados "complejos de ARN_g/Cas9" en este documento) fueron producidos por transcripción *in vitro* (por ejemplo, T7 Quick High Yield RNA Synthesis Kit, New England Biolabs, Ipswich, MA) a partir de moldes de ADN de cadena doble que incorporan un promotor T7 en el extremo 5' de la secuencia de ADN. La reacción en cadena de la polimerasa (PCR) con 5 cebadores superpuestos ensamblaron los moldes de ADN de cadena doble para los componentes de ARN_g. Los oligonucleótidos utilizados en el ensamblaje se presentan en la Tabla 6.

50 Tabla 6: Cebadores superpuestos para la generación de moldes de componentes de ARN_g de Cas9D10A y nucleasa Cas9

Componente	Objetivo para unión de ADN	Cebadores
ARNgu de Cas9 y Cas9D10A	CD34 Objetivo 1	A, B, C, D, E
ARNgu de Cas9 y Cas9D10A	CD34 Objetivo 2	A, B, C, D, F
ARNgu de Cas9 y Cas9D10A	CD34 Objetivo 3	A, B, C, D, G
ARNgu de Cas9 y Cas9D10A	CD34 Objetivo 4	A, B, C, D, H,
ARNgu de Cas9 y Cas9D10A	CD34 Objetivo 5	A, B, C, D, I
A	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC (SEQ ID NO: 3)	
B	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC (SEQ ID NO: 4)	
C	AAAAAAGCACCGACTCGGTGCC (SEQ ID NO: 1)	
D	AGTAATAATACGACTCACTATAG (SEQ ID NO: 2)	
E	TAATACGACTCACTATAGGAACACTGTGCTGATTACAGGTTTTAGAGCTAGAAATAGC (SEQ ID NO: 34)	
F	TAATACGACTCACTATAGGTTTGTGTTTCCATAAACTGGTTTTAGAGCTAGAAATAGC (SEQ ID NO: 35)	
G	TAATACGACTCACTATAGGCTACTAAGTGGCTCCCGTTTTAGAGCTAGAAATAGC (SEQ ID NO: 36)	
H	TAATACGACTCACTATAGTCCCAAAGCGGAGGGCGTTGTTTTAGAGCTAGAAATAGC (SEQ ID NO: 37)	
I	TAATACGACTCACTATAGAGGCTGGGTTGCCGCCGTCGGTTTTAGAGCTAGAAATAGC (SEQ ID NO: 38)	

Los cebadores de ADN estaban presentes en una concentración de 2 nM cada uno. Se usaron dos cebadores de ADN externos correspondientes al promotor T7 (cebador directo: oligonucleótido A, tabla 1) y el extremo 3' de la secuencia de ARN (cebadores inversos: oligonucleótidos C, tabla 1) a 640 nM para impulsar la reacción de amplificación. Las reacciones de PCR se realizaron utilizando el kit Kapa HiFiHotstart PCR^{MR} (Kapa Biosystems, Inc., Wilmington, MA) de acuerdo con las recomendaciones del fabricante. Las reacciones de ensamblaje de PCR se llevaron a cabo utilizando las siguientes condiciones de ciclos térmicos: 98 °C durante 2 minutos, 35 ciclos de 15 segundos a 98 °C, 15 segundos a 62 °C, 15 segundos a 72 °C y una extensión final a 72 °C durante 2 min. Después de la reacción PCR, la cantidad de producto de PCR se determinó utilizando electroforesis capilar en un Analizador de Fragmentos (Advanced Analytical Technologies, Inc., Ames, IA).

Se transcribieron entre 0,25 y 0,5 µg del molde de ADN para los componentes de ARNgu usando el kit de síntesis de ARN de alto rendimiento T7 (New England Biolabs, Ipswich, MA) durante aproximadamente 16 horas a 37°C. Las reacciones de transcripción fueron tratadas con ADNasa (New England Biolabs, Ipswich, MA). La calidad del ARN transcrito se verificó mediante electroforesis capilar en un analizador de fragmentos (Advanced Analytical Technologies, Inc., Ames, IA).

Los componentes proteicos de las RNP se expresaron a partir de vectores de expresión bacterianos en *E. coli* (BL21 (DE3)) y purificado mediante cromatografía de afinidad, intercambio iónico y exclusión por tamaño según los procedimientos descritos en Jinek, M., et al., "A programable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* 337 (2012) 816-821. Este procedimiento para la producción de RNP Cas9 y/o Cas9D10A / Cas9H840A se puede aplicar a la producción de otros RNP Cas9 y/o Cas9D10A/Cas9H840A como se describe en este documento. La secuencia de codificación para Cas9 de *S. pyogenes* incluyó las dos secuencias de localización nuclear (NLS) en el extremo terminal C. Se prepararon variantes de nickasa Cas9D10A o Cas9H840A de Cas9 etiquetada con NLS, en las que se mutó un residuo de sitio activo de cualquier dominio de nucleasa (Jinek, et al., 2012) introduciendo mutaciones en la secuencia codificante de Cas9 de *S. pyogenes* por mutagénesis dirigida al sitio (por ejemplo, kit de mutagénesis dirigida al sitio Q5, New England Biolabs, Ipswich, MA).

Ejemplo 7

Análisis de secuenciación profunda para la detección de modificaciones objetivo en células eucariotas

Este ejemplo ilustra el uso de un secuenciador MiSeq (Illumina, San Diego, CA) para el análisis de secuenciación profunda para cuantificar los eventos de edición totales iniciados por la escisión del ADN por Cas9 o Cas9D10A y comparar los tipos de reparación de ADN. Los ejemplos de tipos de reparación de ADN pueden incluir vías mutagénicas de unión terminal, como la unión terminal no homóloga (NHEJ) o la inserción de material de una

secuencia donante mediante reparación dirigida por homología (HDR). En este ejemplo, Cas9 y Cas9D10A se dirigieron al gen humano CD34 en cinco sitios independientes mediante ARN_g específicos.

A. Transfección de las RNP Cas9 / Cas9D10A:

5 Para ensamblar las RNP Cas9 y Cas9D10A, se incubaron 1,36 µl de ARN_g (que corresponde a aproximadamente 1-5 µg) durante 2 minutos a 95 °C y luego se dejó equilibrar a temperatura ambiente durante aproximadamente 5 minutos. Posteriormente, se mezclaron Cas9 y Cas9D10A con un ARN_g correspondiente para formar RNP en un tampón de reacción (HEPES 20 mM, pH 7,5, KCl 100 mM, MgCl₂ 5 mM, glicerol al 5%). Se combinaron 20 pmoles de Cas9 o Cas9D10A con el ARN_g objetivo y se ensamblaron las RNP funcionales mediante incubación a 37 °C durante 10 minutos. Finalmente, se combinaron 20 pmoles de RNP Cas9 o Cas9D10A con 100 pmoles de molde de oligonucleótido donante de ADN para HDR antes de la transfección en células. Los experimentos se realizaron por triplicado.

Tabla 7: Moldes de donantes de oligonucleótidos de ADN

CD34 Objetivo 1	TCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCACAGAgaaattcCTGTAATC AGCACAGTGTTCACCACCCCAGCCAACGTTTCAACT (SEQ ID NO:39)
CD34 Objetivo 2	CCAGAAACGACAGTCAAATTCACATCTACCTCTGTGATAAgaaattcCAGTTTAT GGAAACACAAACTCTTCTGTCCAGTCACAGACCTCT (SEQ ID NO:40)
CD34 Objetivo 3	ACCCAGCCTCCCTCCTAACGCCCTCCGCCTTTGGGACCAAgaaattcGGGGAGCT CAAGTTAGTAGCAGCCAAGGAGAGGCGCTGCCTTGC (SEQ ID NO:41)
CD34 Objetivo 4	CCACCTTTTTTGGCCTCGACGGCGGCAACCCAGCCTCCCTgaaattcAACGCCCTC CGCCTTTGGGACCAACCAGGGGAGCTCAAGTTAGT (SEQ ID NO:42)
CD34 Objetivo 4	CGAGGCATCTGGAGCCCGAACAAACCTCCACCTTTTTTGGgaaattcCGACGGCG GCAACCCAGCCTCCCTCCTAACGCCCTCCGCCTTTG (SEQ ID NO:43)
CD34 Objetivos 1 + 2 Cas9D10A	CACATCTACCTCTGTGATAAgCTCAGTTTATGGAAAtCACAAACTCTTCTGTCC AGTCACAGAgCTCTGTAATCAGCACAGTGTTCACCA (SEQ ID NO:44)
CD34 Objetivos 3 + 4 Cas9D10A	CCTCGACGGCGGCAACCCAGCCTCCCTgCTAACGCCCTCCGaaTTcTGGGACC AAgCAGGGGAGCTCAAGTTAGTAGCAGCCAAGGAGAG (SEQ ID NO:45)
CD34 Objetivos 4 + 5 Cas9D10A	CCGAACAAACCTCCACCTTTTTTGGGgCTCGACGGCGGCAACCGAattCCTCCCT CgTAACGCCCTCCGCCTTTGGGACCAACCAGGGGAG (SEQ ID NO:46)
CD34 Objetivos 3 + 5 Cas9D10A	CCACCTTTTTTGGgCTCGACGGCGGCAACCCAGCCTCCCTCCgAAAtCGCCCTC CGCCTTTGGGACCAAgCAGGGGAGCTCAAGTTAGTA (SEQ ID NO:47)

15 Los complejos de Cas9/Cas9D10ARNP se transfectaron en células K562 (ATCC, Manassas, VA), utilizando el sistema Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ) y el siguiente protocolo: RNP y RNP más complejos donantes se dispensaron en un volumen final de 2-3 µl en pocillos individuales de una placa de 96 pocillos. Las células K562 suspendidas en medio se transfirieron del matraz de cultivo a un tubo cónico de 50 ml, luego las células se sedimentaron por centrifugación durante 3 minutos a 200 x g, el medio de cultivo se aspiró y se lavó una vez con PBS libre de calcio y magnesio. Las células K562 se sedimentaron luego por centrifugación durante 3 minutos a 200 x g, se aspiró el PBS y el sedimento celular se resuspendió en 10 ml de PBS libre de calcio y magnesio.

25 Las células K562 se contaron utilizando el Countess® II Automated Cell Counter^{MR} (Life Technologies, Grand Island, NY). Se transfirieron 2,2 x 10⁷ células a un tubo de 50 ml y se sedimentaron. El PBS se aspiró y las células se resuspendieron en una solución de Nucleofector^{MR} SF (Lonza, Allendale, NJ) a una densidad de 1 x 10⁷ células/ml. Luego se agregan 20 µl de la suspensión celular a los pocillos individuales que contienen 2-3 µl de RNP y RNP más complejos donantes y todo el volumen se transfirió a los pocillos de una placa Nucleocuvette^{MR} de 96 pocillos (Lonza, Allendale, NJ). La placa se cargó en el Nucleofector^{MR} 96-well Shuttle^{MR} (Lonza, Allendale, NJ) y las células se nucleofectaron utilizando el programa 96-FF-120 Nucleofector^{MR} (Lonza, Allendale, NJ). Después de la nucleofección, se agregaron 80 µl del medio de Dulbecco modificados de Iscove (IMDM, Life Technologies, Grand Island, NY), complementado con FBS al 10% (Fisher Scientific, Pittsburgh, PA) y complementado con penicilina y estreptomina (Life Technologies, Grand Island, NY), a cada pocillo y se transfirieron 50 µl de la suspensión celular a una placa de cultivo celular de 96 pocillos que contenía 150 µl de medio de cultivo completo IMDM precalentado. La placa se transfirió luego a una incubadora de cultivo de tejidos y se mantuvo a 37 °C en CO₂ al 5% durante aproximadamente 48 horas.

Se aisló ADN genómico (ADNg) de células K562 48 horas después de la transfección de Cas9/Cas9D10A utilizando 50 µl de solución de extracción de ADN QuickExtract (Epicenter, Madison, WI) por pocillo, seguido de incubación a 37 °C durante 10 minutos. Se agregaron 50 µl de agua a las muestras, y luego se incubaron a 75 °C durante 10 minutos y 95 °C durante 5 minutos para detener la reacción. El ADNg se almacenó a -80 °C hasta su posterior procesamiento.

B. Preparación de la biblioteca de secuenciación:

Usando ADNg previamente aislado, se realizó una primera PCR usando Q5 Hot Start High-Fidelity 2X Master Mix^{MR} (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 µM cada uno, 3,75 µl de ADNg en un volumen de 10 L y amplificado a 98 °C durante 1 minuto, 35 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 minutos. La reacción de PCR se diluyó 1:100 en agua. Los cebadores específicos del objetivo se muestran en la Tabla 8.

Tabla 8: Cebadores específicos de objetivo utilizados

CD34 Objetivo 1_F	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCAAGGCTAGTGCTAGTGG (SEQ ID NO: 48)
CD34 Objetivo 1_R	CACTCTTTCCCTACACGACGCTCTTCCGATCTACATGCACACCCATGTTTTG (SEQ ID NO: 49)
CD34 Objetivo 2_F	GGAGTTCAGACGTGTGCTCTTCCGATCTAACATTTCCAGGTGACAGGC (SEQ ID NO: 50)
CD34 Objetivo 2_R	CACTCTTTCCCTACACGACGCTCTTCCGATCTACATGCACACCCATGTTTTG (SEQ ID NO: 51)
CD34 Objetivo 3_F	GGAGTTCAGACGTGTGCTCTTCCGA TCTGTGGGGA TTCTTGCTTTTT (SEQ ID NO: 52)
CD34 Objetivo 3_R	CACTCTTTCCCT ACACGACGCTCTTCCGA TCTCTCCAGAAAGCTGAACGAGG (SEQ ID NO: 53)
CD34 Objetivo 4_F	GGAGTTCAGACGTGTGCTCTTCCGATCTTTTCTCTCTTCTCCCCTCC (SEQ ID NO: 54)
CD34 Objetivo 4_R	CACTCTTTCCCT ACACGACGCTCTTCCGA TCTCTGCCACAAAGGGGTT AAAA (SEQ ID NO: 55)
CD34 Objetivo 5_F	GGAGTTCAGACGTGTGCTCTTCCGATCTTTTCTCTCTTCTCCCCTCC (SEQ ID NO: 56)
CD34 Objetivo 5_R	CACTCTTTCCCT ACACGACGCTCTTCCGA TCTCTGCCACAAAGGGGTT AAAA (SEQ ID NO: 57)
CD34 Objetivos 1+2 Cas9D10A_F	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCAAGGCTAGTGCTAGTGG (SEQ ID NO: 58)
CD34 Objetivos 1+2 Cas9D10A_R	CACTCTTTCCCTACACGACGCTCTTCCGATCTCACATGCACACCCATGTTTT (SEQ ID NO: 59)
CD34 Objetivos 3-5 Cas9D10A_F	GGAGTTCAGACGTGTGCTCTTCCGATCTTCTCTTCTCCCCTCCCTTTT (SEQ ID NO: 60)
CD34 Objetivos 3-5 Cas9D10A_R	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCCACAAAGGGGTTAAAAGTT (SEQ ID NO: 61)

Se configuró una segunda PCR de "código de barras" utilizando cebadores únicos para cada muestra que facilita la secuenciación múltiplex. Los pares de cebadores se muestran en la Tabla 9.

Tabla 9: Cebadores de códigos de barras

ILMN_AMP_FORi5_BC9	AATGATACGGCGACCACCGAGATCTACACTGAACCTTACACTCTTCCCTACACGACG (SEQ ID NO: 62)
--------------------	---

(continuación)

ILMN_AMP_FORi5_BC10	AATGATACGGCGACCACCGAGATCTACACTGCTAAGTACACTCTTTCCTACACGACG (SEQ ID NO: 63)
ILMN_AMP_FORi5_BC11	AATGATACGGCGACCACCGAGATCTACACTAAGTTCACACTCTTTCCTACACGACG (SEQ ID NO: 64)
ILMN_AMP_FORi5_BC12	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCTACACGACG (SEQ ID NO: 65)
ILMN_AMP_FORi5_BC13	AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCTACACGACG (SEQ ID NO: 66)
ILMN_AMP_FORi5_BC14	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCTACACGACG (SEQ ID NO: 67)
ILMN_AMP_FORi5_BC15	AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCTACACGACG (SEQ ID NO: 68)
ILMN_AMP_FORi5_BC16	AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCTACACGACG (SEQ ID NO: 69)
ILMN_AMP_REVi7_BC49	CAAGCAGAAGACGGCATAACGAGATATACATCGGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 70)
ILMN_AMP_REVi7_BC50	CAAGCAGAAGACGGCATAACGAGATATGCCTAAGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 71)
ILMN_AMP_REVi7_BC51	CAAGCAGAAGACGGCATAACGAGATATTCAAGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 72)
ILMN_AMP_REVi7_BC52	CAAGCAGAAGACGGCATAACGAGATATCTGATCGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 73)
ILMN_AMP_REVi7_BC53	CAAGCAGAAGACGGCATAACGAGATATGTAGCCGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 74)
ILMN_AMP_REVi7_BC54	CAAGCAGAAGACGGCATAACGAGATATTTGACTGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 75)
ILMN_AMP_REVi7_BC55	CAAGCAGAAGACGGCATAACGAGATATGGAAGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 76)
ILMN_AMP_REVi7_BC56	CAAGCAGAAGACGGCATAACGAGATATTGACATGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 77)
ILMN_AMP_REVi7_BC57	CAAGCAGAAGACGGCATAACGAGATATGGACGGGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 78)
ILMN_AMP_REVi7_BC58	CAAGCAGAAGACGGCATAACGAGATATCCACTCGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 79)
ILMN_AMP_REVi7_BC59	CAAGCAGAAGACGGCATAACGAGATATCTTTTGGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 80)
ILMN_AMP_REVi7_BC60	CAAGCAGAAGACGGCATAACGAGATATTGAGTGGTGACTGGAGTTCAGACGTGTGCTC (SEQ ID NO: 81)

5 La segunda PCR se realizó utilizando Q5 Hot Start High-Fidelity 2X Master Mix^{MR} (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 µM cada uno, 1 µl de una PCR diluida primero 1:100, en un volumen final de 10 µl y se amplificó a 98 °C durante 1 minuto, 12 ciclos de 10 s a 98 °C, 20 s a 60 °C, 30 s a 72 °C y una extensión final a 72 °C durante 2 minutos. Las reacciones de PCR se agruparon en un solo tubo de microcentrifuga para la limpieza de amplicones para secuenciación basada en perlas SPRIselect^{MR} (Beckman Coulter, Pasadena,

CA).

5 A los amplicones reunidos, se les agregaron 0,9 volúmenes de perlas SPRIselect^{MR}, se mezclaron y se incubaron a temperatura ambiente (RT) durante 10 minutos. El tubo de microcentrifuga se colocó en un soporte de tubo magnético (Beckman Coulter, Pasadena, CA) hasta que se aclaró la solución. El sobrenadante se eliminó y se desechó, y las perlas residuales se lavaron con 1 volumen de etanol al 85% y se incubaron a temperatura ambiente durante 30 s. Después de la incubación, se aspiró el etanol y las perlas se secaron al aire a temperatura ambiente durante 10 min. Luego se retiró el tubo de microcentrifuga del soporte magnético y se agregaron a las perlas 0,25x volúmenes de agua (Qiagen, Venlo, Limburg), se mezclaron vigorosamente y se incubaron durante 2 minutos a temperatura ambiente. El tubo de microcentrifuga se centrifugó en una microcentrifuga para recoger el contenido del tubo y luego se devolvió al imán, se incubó hasta que la solución se aclaró y el sobrenadante que contenía los amplicones purificados se dispensó en un tubo limpio de microcentrifuga. La biblioteca de amplicones purificada se cuantificó utilizando el sistema Nanodrop^{MR} 2000 (Thermo Scientific, Wilmington DE).

15 La biblioteca de amplicones se normalizó a una concentración de 4 nmol calculada a partir de los valores de Nanodrop y el tamaño de los amplicones. La biblioteca se analizó en el secuenciador MiSeq con el MiSeq Reagent Kit v2^{MR}, 300 Ciclos (Illumina, San Diego), con dos procesos de finalización pareados de 151 ciclos más dos lecturas de índice de ocho ciclos.

C. Análisis de datos de secuenciación profunda:

20 La identidad de los productos en los datos de secuenciación se analizó con base en la indexación de secuencias de código de barras adaptadas a los amplicones en la segunda ronda de PCR. Se utilizó un guión computacional para procesar los datos de MiSeq ejecutando las siguientes tareas:

1. Las lecturas se alinearon con el genoma humano (compilación GRCh38/38) utilizando el software Bowtie (bowtie-bio.sourceforge.net/index.shtml).
2. Las lecturas alineadas se compararon con los loci de tipo silvestre. Las lecturas que no se alineaban con ninguna parte de los loci fueron descartadas.
- 25 3. Las lecturas que coinciden con la secuencia de tipo silvestre se registraron. Las lecturas con inserciones o supresiones (alrededor de 5 pb del sitio de corte de Cas9) se clasificaron por el tipo de inserciones o supresiones y se contabilizaron.
4. Las lecturas de inserciones o supresiones totales se dividieron por la suma de las lecturas de tipo silvestre y las lecturas de inserciones o supresiones para obtener lecturas del porcentaje mutado.

30 Las estructuras de inserciones y supresiones se compararon entre los datos de secuencia que se generaron a partir de células transfectadas con RNP Cas9 natural o RNP Cas9 + Donante, para cada uno de los objetivos individuales, y para RNP Cas9D10A y RNP Cas9D10A + Donante para cada uno de los pares objetivos. Los datos experimentales demostraron que las células transfectadas con RNP Cas9 exhibieron una serie de clases de ediciones mutantes. RNP Cas9 + Donante mostró un espectro similar de ediciones mutantes y ediciones dependientes del donante, mientras que las células transfectadas con RNP Cas9D10A solamente, no mostraron evidencia de edición, pero RNP Cas9D10A + Donante demostró niveles similares de inserción del donante en la RNP Cas9 + Donante, pero sin ediciones mutantes medibles que no puedan atribuirse a la incorporación de la secuencia del donante.

40 La Fig. 12 muestra una comparación de tipos de reparación utilizando Cas9 o Cas9D10A en los Objetivos 3 y 4 (locus CD34 humano). RNP Cas9 complejada con ARN_g se usó para apuntar a CD34 Objetivo 3 o CD34 Objetivo 4. Las RNP Cas9D10A complejadas con ARN_g se usaron para apuntar a CD34 Objetivo 3 y Objetivo 4. Los controles negativos fueron Cas9 o Cas9D10A solamente, sin ARN_g. La distribución de la reparación se muestra por las barras. Como puede verse, RNP Cas9 mostró solo reparación mutagénica. RNP Cas9 + Donante demostró reparación mutagénica y HDR, mientras que RNP Cas9D10A mostró reparación mutagénica apenas detectable. RNP Cas9D10A + Donante demostró ediciones de HDR con reparación mutagénica apenas detectable.

45 La Tabla 10 contiene un promedio de tres repeticiones (excluyendo los controles negativos n = 2) y la desviación estándar (STD) de cada clase.

Tabla 10: Datos utilizados en la Fig. 12

Muestra	Nucleasa	% No editado	% de reparación mutagénica	% de HDR	STD no editada	STD MUT	STD de HDR
Objetivo 3	Cas9	7,7	92,3	0	0,57	0,57	0

(continuación)

Muestra	Nucleasa	% No editado	% de reparación mutagénica	% de HDR	STD no editada	STD MUT	STD de HDR
Objetivo 4	Cas9	58,3	41,7	0	2,39	2,3	0
Objetivo 3+Donante	Cas9	2	59,7	38,3	0	2,31	2,31
Objetivo 4+Donante	Cas9	22,3	54,3	23,3	2,08	2,51	1,52
Objetivo 3 neg	Cas9	100	0	0	ND	ND	ND
Objetivo 4 neg	Cas9	100	0	0	ND	ND	ND
Objetivo 3+4	Cas9D10A	99,3	0,7	0	0,05	0,05	0
Objetivo 3+4 Donante	Cas9D10A	81,8	0,5	17,6	1,52	0	1,52
Objetivo 3+4 neg	Cas9D10A	99,4	0,6	0	ND	ND	ND

5 Siguiendo los lineamientos de la presente memoria descriptiva y los ejemplos, el experto en la técnica puede practicar el análisis de secuenciación profunda descrito en este ejemplo con otros complejos de RNP Cas9/Cas9D10A (es decir, ensamblados con diferentes ARN_gu y distintas proporciones de Cas9/Cas9D10A y moldes de oligonucleótidos donantes.

Ejemplo 8

Uso de nickasas en tándem Cas9D10A emparejadas o Cas9H840A emparejadas para mejorar la proporción de ediciones específicas de HDR en un sitio de rompimiento

10 Este ejemplo ilustra el uso de un mutante de nickasa Cas9 en el que un dominio de nucleasa se inactivará (ya sea Cas9D10A o Cas9H840A) para participar preferentemente en las rutas de HDR y bloquear las rutas de reparación mutagénica en los sitios de rompimiento en las células eucarióticas. En este ejemplo, las secuencias espaciadoras para las dos secuencias de ARN_gu se eligen para variar la duración de la supresión alrededor del sitio objetivo deseado. Las secuencias se eligen de modo que las nickasas emparejadas se dirijan a dos sitios en la misma
15 cadena variando la distancia entre dos sitios en un intervalo de separación de 20 a 2000 nucleótidos. Los polinucleótidos donantes se diseñan con diferentes longitudes y posiciones en relación con las ubicaciones de las secuencias espaciadoras y se prueban en combinación con cada par de RNP_gu de nickasa Cas9. Usando los procedimientos descritos en los Ejemplos 6 y 7, se realizan experimentos para medir la frecuencia y el tipo de
20 reparación de ADN que tiene lugar con cada combinación de nickasas pareadas. Los datos se analizan para identificar la combinación de RNP_gu nickasa y el polinucleótido donante que conduce a la frecuencia más alta de HDR con la frecuencia más baja de edición mutante.

Ejemplo 9

25 Uso de nickasas en tándem Cas9D10A emparejadas o Cas9H840A emparejadas para mejorar la proporción de ediciones específicas de HDR en un sitio de rompimiento e introducir diferentes inserciones o supresiones de nucleótidos específicas

30 Este ejemplo ilustra el uso de un mutante de nickasa Cas9 en el que un dominio de nucleasa se inactivará (ya sea Cas9D10A o Cas9H840A) para participar preferentemente en las rutas de HDR y bloquear las rutas de reparación mutagénica en los sitios de rompimiento en las células eucarióticas. En este ejemplo, tanto Cas9D10A emparejada como Cas9H840A emparejada se usan con dos ARN_gu específicos que suministran las nickasas emparejadas a dos sitios en la misma cadena separados 20-2000 nucleótidos. Los oligonucleótidos donantes están diseñados para administrar inserciones o supresiones de nucleótidos específicos en el sitio deseado (Figura 11). Los experimentos se llevan a cabo con un espaciado variable entre las nickasas y diferentes secuencias donantes y de longitud como se describe en el Ejemplo 8 para identificar la combinación de reactivos que conduce a la frecuencia más alta de HDR y la frecuencia más baja de reparación mutagénica para introducir la modificación deseada en el sitio deseado.

35 Ejemplo 10

Uso de nickasas en tándem Cas9D10A emparejadas o Cas9H840A emparejadas para mejorar la proporción de

ediciones específicas de HDR en un sitio de rompimiento en células primarias humanas con diversas configuraciones donantes

Este ejemplo ilustra el uso de un mutante de nickasa Cas9 en el que se inactivará un dominio de nucleasa (ya sea Cas9D10A o Cas9H840A) para participar exclusivamente en las rutas de HDR y bloquear las rutas de reparación mutagénica en los sitios de rompimiento en células eucarióticas. En este ejemplo, tanto Cas9D10A emparejada como Cas9H840A emparejada pueden usarse en tándem complejadas con dos ARN_g específicos que suministran las nickasas emparejadas a dos sitios en la misma cadena separados 20-2000 nucleótidos. Los oligonucleótidos donantes se proporcionan en diferentes orientaciones y/o longitudes para suministrar inserciones o supresiones de nucleótidos específicos entre dos sitios objetivo de nickasa Cas9 en células primarias humanas para obtener una ventaja terapéutica.

Ejemplo 11

Uso de nickasas en tándem Cas9D10A emparejadas y Cas9H840A para mejorar la proporción de ediciones específicas de HDR en un sitio de rompimiento en células primarias humanas con diversas configuraciones donantes

Este ejemplo ilustra el uso de pares de mutantes de nickasa Cas9 para participar preferentemente en vías de reparación dirigidas por homología y bloquear vías de reparación mutagénica en sitios de rompimiento en células eucariotas. En este ejemplo, Cas9D10A y Cas9H840A se usan en combinación con dos ARN_g específicos que suministran las nickasas emparejadas a dos sitios, lo que da como resultado un corte de la misma cadena separado 20-2000 nucleótidos. Los ARN_g emparejados con Cas9D10A deben elegirse para dirigirse a las secuencias protoespaciadoras y las secuencias PAM en una cadena. Los ARN_g emparejados con Cas9H840A deben elegirse para dirigirse a las secuencias protoespaciadoras y las secuencias PAM en la cadena opuesta a los ARN_g de Cas9D10A para asegurar que la misma cadena se corte dos veces. Los RNP_g se ensamblan por separado para cada mutante de la nickasa combinando la proteína con el ARN_g seleccionado. Los oligonucleótidos donantes están diseñados para brindar inserciones o supresiones específicas de nucleótidos en el sitio deseado (Figura 11) y son sintetizados por un fabricante de oligonucleótidos (por ejemplo, Integrated DNA Technologies, Coralville, IA). Cas9D10a-RNP_g se mezclan con Cas9H840A-RNP_g antes de la transfección y el par de nickasas dirigidas a la misma cadena transfectada junto con el oligonucleótido donante usando los procedimientos descritos en los ejemplos anteriores. Los experimentos se llevan a cabo variando el espaciamiento entre las nickasas y variando la secuencia y la longitud del donante como se describe en el Ejemplo 8 para identificar la combinación de reactivos que conduce a la frecuencia más alta de HDR y la frecuencia más baja de reparación mutagénica para introducir la modificación deseada en el sitio deseado.

Listado de secuencias

<110> CARIBOU BIOSCIENCES, INC.

CAMERON, Peter S.

HAURWITZ, Rachel E.

MAY, Andrew P.

NYE, Christopher H.

OVERBEEK, Megan Van

<120> PROCEDIMIENTOS PARA INCREMENTAR LA EFICIENCIA DE LA MODIFICACIÓN MEDIADA POR CAS9

<130> CBI014.30

<150> US 62/042.358

<151> 2014-08-27

<150> US 62/047.495

<151> 2014-09-08

<160> 84

<170> PatentIn versión 3.5

<210> 1

<211> 23

<212> ADN

<213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 1
 5 aaaaaaagca ccgactcggg gcc 3
 <210> 2
 <211> 23
 <212> ADN
 <213> Secuencia Artificial
 10 <220>
 <223> Oligonucleótido Sintético
 <400> 2
 agtaataata cgactcacta tag 23
 <210> 3
 15 <211> 50
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 20 <400> 3
 gtttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac 50
 <210> 4
 <211> 50
 <212> ADN
 25 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 4
 aaaaaaagca ccgactcggg gccactttt caagtgata acgactagc 50
 30 <210> 5
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 <220>
 35 <223> Oligonucleótido Sintético
 <400> 5
 taatagcact cactataggg tggggggagt ttgctccggt ttagagctag aaatagc 57

- <210> 6
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
- 5 <220>
 <223> Oligonucleótido Sintético
 <400> 6
 taatagcact cactataggg gccactaggg acaggatggt ttagagctag aaatagc 57
 <210> 7
- 10 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
- 15 <400> 7
 taatagcact cactatagtg gagggagttt gctcctgggt ttagagctag aaatagc 57
 <210> 8
 <211> 57
 <212> ADN
- 20 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 8
 taatagcact cactatagga cggatttggt ggatggagtt ttagagctag aaatagc 57
- 25 <210> 9
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 <220>
- 30 <223> Oligonucleótido Sintético
 <400> 9
 taatagcact cactatagca ggacattctg acaccccgtt ttagagctag aaatagc 57
 <210> 10
 <211> 57
- 35 <212> ADN
 <213> Secuencia Artificial
 <220>

ES 2 730 378 T3

<223> Oligonucleótido Sintético
<400> 10
taatagcact cactatagga ggctcccatc acgggggggt ttagagctag aaatagc 57
<210> 11
5 <211> 57
<212> ADN
<213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético
10 <400> 11
taatagcact cactatagtg gggatcacag gttcccgggt ttagagctag aaatagc 57
<210> 12
<211> 57
<212> ADN
15 <213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético
<400> 12
taatagcact cactatagag agctctctg actacaggt ttagagctag aaatagc 57
20 <210> 13
<211> 57
<212> ADN
<213> Secuencia Artificial
<220>
25 <223> Oligonucleótido Sintético
<400> 13
taatagcact cactatagga ccaaatgaga ccagtcgggt ttagagctag aaatagc 57
<210> 14
<211> 57
30 <212> ADN
<213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético
<400> 14
35 taatagcact cactatagcc cattatgata gggaggggt ttagagctag aaatagc 57
<210> 15
<211> 57

<212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 5 <400> 15
 taatagcact cactatagct cctgggatg gaagggcgtt ttagagctag aatagc 57
 <210> 16
 <211> 52
 <212> ADN
 10 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 16
 cactcttcc ctacacgacg ctctccgat ctccagatgg cacattgtca ga 52
 15 <210> 17
 <211> 47
 <212> ADN
 <213> Secuencia Artificial
 <220>
 20 <223> Oligonucleótido Sintético
 <400> 17
 ggagttcaga cgtgtgctct tccgatctcc tagtgactgc cgtctgc 47
 <210> 18
 <211> 48
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 18
 30 ggagttcaga cgtgtgctct tccgatctac ctggccatca tccttcta 48
 <210> 19
 <211> 52
 <212> ADN
 <213> Secuencia Artificial
 35 <220>
 <223> Oligonucleótido Sintético
 <400> 19

	cactctttcc ctacacgacg ctcttccgat ctacagacac ccactgagtc aa	52
	<210> 20	
	<211> 57	
	<212> ADN	
5	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
	<220>	
	<221> característica miscelánea	
10	<222> (25)..(32)	
	<223> n is a, c, g, o t	
	<400> 20	
	caagcagaag acggcatacg agatnnnnn nngtactgg agtcagacg tgtgctc	57
	<210> 21	
15	<211> 58	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
20	<220>	
	<221> característica miscelánea	
	<222> (30)..(37)	
	<223> n is a, c, g, o t	
	<400> 21	
25	aatgatcgg cgaccaccga gatctacacn nnnnnnaca ctcttcct acacgacg	58
	<210> 22	
	<211> 103	
	<212> ARN	
	<213> Secuencia Artificial	
30	<220>	
	<223> Oligonucleótido Sintético	
	<400> 22	
	gggugggggg aguuugcucc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60
	cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu	103
	<210> 23	
35	<211> 103	

ES 2 730 378 T3

<212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 5 <400> 23
 ggggccacua gggacaggau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 24
 <211> 103
 <212> ARN
 10 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 24
 guggagggag uuugcuccug guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 15 <210> 25
 <211> 103
 <212> ARN
 <213> Secuencia Artificial
 <220>
 20 <223> Oligonucleótido Sintético
 <400> 25
 ggacggauuu gugggaugga guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 26
 <211> 104
 25 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 26
 gcaggacauu cugacacccc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 30 <400> 27
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuuu 104
 <210> 27
 <211> 103

ES 2 730 378 T3

<212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 5 <400> 27
 ggaggcuccc aucacggggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 28
 <211> 103
 <212> ARN
 10 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 28
 guggggauca cagguucccc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 15 <210> 29
 <211> 103
 <212> ARN
 <213> Secuencia Artificial
 <220>
 20 <223> Oligonucleótido Sintético
 <400> 29
 gagagcucuu cugacuacag guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 30
 <211> 103
 25 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 30
 ggacccaaug agaccagucc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 30 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 31
 <211> 103

ES 2 730 378 T3

<212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 5 <400> 31
 gcccauuaug auagggaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103
 <210> 32
 <211> 104
 <212> ARN
 10 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 32
 gcuccugggg auuggaaggg cguuuuagag cuagaaauag caaguuaaaau uaaggcuagu 60
 ccguuaucaa cuugaaaaag uggcaccgag ucgugcuuuu uuuu 104
 15 <210> 33
 <211> 8
 <212> ADN
 <213> Secuencia Artificial
 <220>
 20 <223> Oligonucleótido Sintético
 <220>
 <221> característica miscelánea
 <222> (1)..(8)
 <223> n is a, c, g, o t
 25 <400> 33
 nnnnnnnn 8
 <210> 34
 <211> 58
 <212> ADN
 30 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 34
 taatagcact cactatagga acactgtgct gattacaggt ttagagcta gaaatagc 58

- <210> 35
 <211> 58
 <212> ADN
 <213> Secuencia Artificial
- 5 <220>
 <223> Oligonucleótido Sintético
 <400> 35
 taatagcact cactataggt ttgtgttcc ataaactggt ttagagcta gaaatagc 58
 <210> 36
- 10 <211> 58
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
- 15 <400> 36
 taatagcact cactataggc tactaactg agctccccgt ttagagcta gaaatagc 58
 <210> 37
 <211> 58
 <212> ADN
- 20 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 37
 taatagcact cactatagtc ccaaaggcgg agggcggtgt ttagagcta gaaatagc 58
- 25 <210> 38
 <211> 58
 <212> ADN
 <213> Secuencia Artificial
 <220>
- 30 <223> Oligonucleótido Sintético
 <400> 38
 taatagcact cactatagag gctgggtgc cgccgtcgg ttagagcta gaaatagc 58
 <210> 39
 <211> 90
- 35 <212> ADN
 <213> Secuencia Artificial
 <220>

ES 2 730 378 T3

<223> Oligonucleótido Sintético
 <400> 39
 tcagtttatg gaaacacaaa ctcttctgtc cagtcacaga gaattcctgt aatcagcaca 60
 gtgttcacca ccccagccaa cgtttcaact 90
 <210> 40
 5 <211> 90
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 10 <400> 40
 ccagaaacga cagtcaaatt cacatctacc tctgtgataa gaattccagt ttatggaaac 60
 acaaactctt ctgtccagtc acagacctct 90
 <210> 41
 <211> 90
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 41
 accagcctc cctcctaacg ccctccgctt ttgggaccaa gaattcgggg agctcaagtt 60
 agtagcagcc aaggagaggc gctgccttgc 90
 20 <210> 42
 <211> 90
 <212> ADN
 <213> Secuencia Artificial
 <220>
 25 <223> Oligonucleótido Sintético
 <400> 42
 ccaccttttt tggcctcgac ggcggcaacc cagcctccct gaattcaacg ccctccgctt 60
 ttgggaccaa ccaggggagc tcaagttagt 90
 <210> 43
 <211> 90
 30 <212> ADN
 <213> Secuencia Artificial
 <220>

ES 2 730 378 T3

<223> Oligonucleótido Sintético
 <400> 43
 cgaggcatct ggagcccga caaacctcca ccttttttgg gaattccgac ggcggcaacc 60
 cagcctccct cctaacgccc tccgcctttg 90
 <210> 44
 5 <211> 90
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 10 <400> 44
 cacatctacc tctgtgataa gctcagttta tggaaattcac aaactcttct gtccagtcac 60
 agagctctgt aatcagcaca gtgttcacca 90
 <210> 45
 <211> 90
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 45
 cctcgacggc ggcaaccag cctccctgct aacgccctcc gaattctggg accaagcagg 60
 ggagctcaag ttagtagcag ccaaggagag 90
 20 <210> 46
 <211> 90
 <212> ADN
 <213> Secuencia Artificial
 <220>
 25 <223> Oligonucleótido Sintético
 <400> 46
 ccgaacaaac ctccaccttt tttggcgtcg acggcggcaa ccgaattcct ccctcgtaac 60
 gccctccgcc tttgggacca accaggggag 90
 <210> 47
 <211> 90
 30 <212> ADN
 <213> Secuencia Artificial
 <220>

ES 2 730 378 T3

<223> Oligonucleótido Sintético
 <400> 47
 ccaccttttt tgggctcgac ggcggcaacc cagcctccct ccgaattcgc cctccgcctt 60
 tgggaccaag caggggagct caagttagta 90
 <210> 48
 5 <211> 48
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 10 <400> 48
 ggagtcaga cgtgtgctct tccgatctg caaggctagt gctagtgg 48
 <210> 49
 <211> 52
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 49
 cactcttcc ctacagcag ctctccgat ctacatgcac acctatgttt tg 52
 20 <210> 50
 <211> 48
 <212> ADN
 <213> Secuencia Artificial
 <220>
 25 <223> Oligonucleótido Sintético
 <400> 50
 ggagtcaga cgtgtgctct tccgatctaa cattccagg tgacaggc 48
 <210> 51
 <211> 52
 30 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 51
 35 cactcttcc ctacagcag ctctccgat ctacatgcac acctatgttt tg 52

- <210> 52
 <211> 48
 <212> ADN
 <213> Secuencia Artificial
- 5 <220>
 <223> Oligonucleótido Sintético
 <400> 52
 ggagttcaga cgtgtgctct tccgatctgt gggggattct tgctttt 48
 <210> 53
- 10 <211> 52
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
- 15 <400> 53
 cactctttcc ctacacgacg ctctccgat ctctccagaa agctgaacga gg 52
 <210> 54
 <211> 48
 <212> ADN
- 20 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 54
 ggagttcaga cgtgtgctct tccgatcttt tcctctcttc tcccctcc 48
- 25 <210> 55
 <211> 52
 <212> ADN
 <213> Secuencia Artificial
 <220>
- 30 <223> Oligonucleótido Sintético
 <400> 55
 cactctttcc ctacacgacg ctctccgat ctctgccaca aaggggtaa aa 52
 <210> 56
 <211> 48
- 35 <212> ADN
 <213> Secuencia Artificial
 <220>

<223> Oligonucleótido Sintético
 <400> 56
 ggagttcaga cgtgtgctct tccgatctt tctctcttc tcccctcc 48
 <210> 57
 5 <211> 52
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 10 <400> 57
 cactcttcc ctacacgacg ctctccgat ctctgccaca aagggttaa aa 52
 <210> 58
 <211> 48
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 58
 ggagttcaga cgtgtgctct tccgatcttg caaggctagt gctagtgg 48
 20 <210> 59
 <211> 52
 <212> ADN
 <213> Secuencia Artificial
 <220>
 25 <223> Oligonucleótido Sintético
 <400> 59
 cactcttcc ctacacgacg ctctccgat ctacatgca cacccatgtt tt 52
 <210> 60
 <211> 48
 30 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 60
 35 ggagttcaga cgtgtgctct tccgatcttc tcttctccc tccctttt 48
 <210> 61
 <211> 53

ES 2 730 378 T3

<212> ADN
<213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético

5 <400> 61
cactctttcc ctacacgacg ctctccgat ctgccacaaa ggggttaaaa gtt 53
<210> 62
<211> 58
<212> ADN

10 <213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético
<400> 62
aatgatacgg cgaccaccga gatctacact gaaccttaca ctcttcct acacgacg 58

15 <210> 63
<211> 58
<212> ADN
<213> Secuencia Artificial
<220>

20 <223> Oligonucleótido Sintético
<400> 63
aatgatacgg cgaccaccga gatctacact gctaagtaca ctcttcct acacgacg 58
<210> 64
<211> 58

25 <212> ADN
<213> Secuencia Artificial
<220>
<223> Oligonucleótido Sintético
<400> 64

30 aatgatacgg cgaccaccga gatctacact aagtccaca ctcttcct acacgacg 58
<210> 65
<211> 58
<212> ADN
<213> Secuencia Artificial

35 <220>
<223> Oligonucleótido Sintético
<400> 65

	aatgatacgg cgaccaccga gatctacaca tagaggcaca ctcttcct acacgacg	58
	<210> 66	
	<211> 58	
	<212> ADN	
5	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
	<400> 66	
	aatgatacgg cgaccaccga gatctacacg gctctgaaca ctcttcct acacgacg	58
10	<210> 67	
	<211> 58	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
15	<223> Oligonucleótido Sintético	
	<400> 67	
	aatgatacgg cgaccaccga gatctacaca ggccaagaca ctcttcct acacgacg	58
	<210> 68	
	<211> 58	
20	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
	<400> 68	
25	aatgatacgg cgaccaccga gatctacact aatcttaaca ctcttcct acacgacg	58
	<210> 69	
	<211> 58	
	<212> ADN	
	<213> Secuencia Artificial	
30	<220>	
	<223> Oligonucleótido Sintético	
	<400> 69	
	aatgatacgg cgaccaccga gatctacacc aggacgtaca ctcttcct acacgacg	58
	<210> 70	
35	<211> 57	
	<212> ADN	
	<213> Secuencia Artificial	

ES 2 730 378 T3

	<220>	
	<223> Oligonucleótido Sintético	
	<400> 70	
	caagcagaag acggcatacg agatatacat cgggactgg agttcagacg tggctc	57
5	<210> 71	
	<211> 57	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
10	<223> Oligonucleótido Sintético	
	<400> 71	
	caagcagaag acggcatacg agatagcct aagtgactgg agttcagacg tggctc	57
	<210> 72	
	<211> 57	
15	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
	<400> 72	
20	caagcagaag acggcatacg agatattcaa gttgactgg agttcagacg tggctc	57
	<210> 73	
	<211> 57	
	<212> ADN	
	<213> Secuencia Artificial	
25	<220>	
	<223> Oligonucleótido Sintético	
	<400> 73	
	caagcagaag acggcatacg agatatctga tgggactgg agttcagacg tggctc	57
	<210> 74	
30	<211> 57	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> Oligonucleótido Sintético	
35	<400> 74	
	caagcagaag acggcatacg agatattgag cgggactgg agttcagacg tggctc	57
	<210> 75	

- <211> 57
 <212> ADN
 <213> Secuencia Artificial
 <220>
- 5 <223> Oligonucleótido Sintético
 <400> 75
 caagcagaag acggcatacg agatattga ctgtgactgg agttcagacg tgtgctc 57
 <210> 76
 <211> 57
- 10 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 76
- 15 caagcagaag acggcatacg agatattgaa ctgtgactgg agttcagacg tgtgctc 57
 <210> 77
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
- 20 <220>
 <223> Oligonucleótido Sintético
 <400> 77
 caagcagaag acggcatacg agatattgac atgtgactgg agttcagacg tgtgctc 57
 <210> 78
- 25 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
- 30 <400> 78
 caagcagaag acggcatacg agatattgac gggtgactgg agttcagacg tgtgctc 57
 <210> 79
 <211> 57
 <212> ADN
- 35 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético

ES 2 730 378 T3

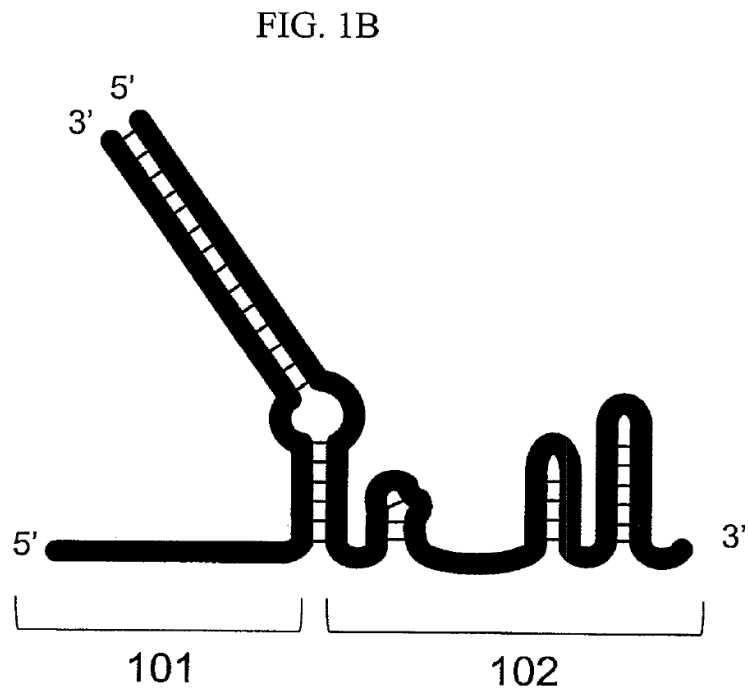
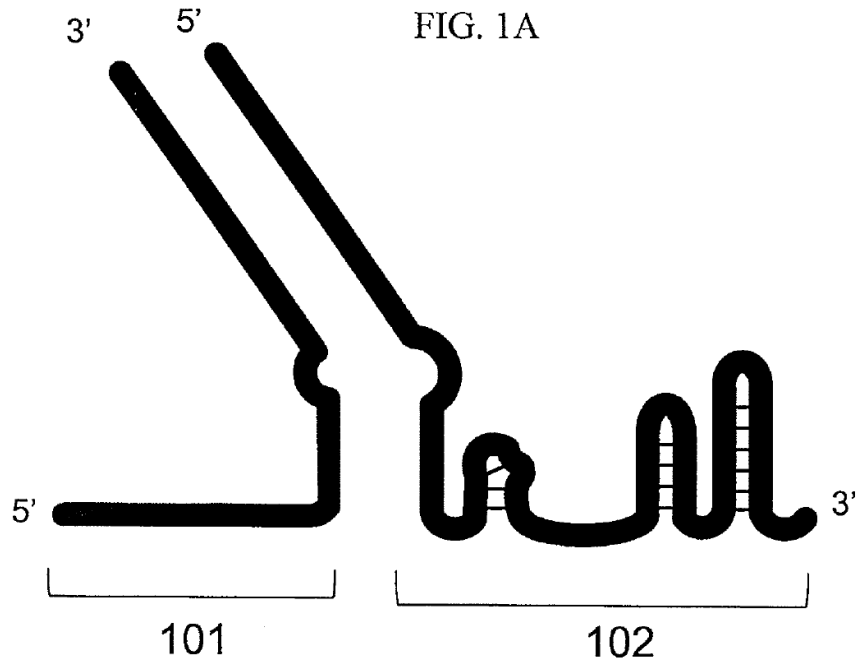
<400> 79
 caagcagaag acggcatacg agatatccac tcgtgactgg agttcagacg tgtgctc 57
 <210> 80
 <211> 57
 5 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido Sintético
 <400> 80
 10 caagcagaag acggcatacg agatatcttt tggtagactgg agttcagacg tgtgctc 57
 <210> 81
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 15 <220>
 <223> Oligonucleótido Sintético
 <400> 81
 caagcagaag acggcatacg agatattgag tggtagactgg agttcagacg tgtgctc 57
 <210> 82
 20 <211> 23
 <212> ADN
 <213> Homo sapiens
 <400> 82
 gggtagggggg agtttgctcc tgg 23
 25 <210> 83
 <211> 23
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> Oligonucleótido Sintético
 <400> 83
 ggatggaggg agtttgctcc tgg 23
 <210> 84
 <211> 23
 35 <212> ADN
 <213> Virus Adeno-Asociado
 <400> 84
 gggccacta gggacaggat tgg 23

REIVINDICACIONES

1. Un procedimiento para reducir la unión y/o la escisión de un ácido nucleico alejado del objetivo por un primer complejo que comprende una proteína Cas9 catalíticamente activa y un primer polinucleótido guía, comprendiendo el procedimiento:
 - 5 poner en contacto el primer complejo con un ácido nucleico objetivo seleccionado, en el que el primer polinucleótido guía comprende un espaciador adaptado para unirse al ácido nucleico objetivo seleccionado; y
 - 10 poner en contacto un segundo complejo con el ácido nucleico alejado del objetivo, en el que el segundo complejo comprende (i) una proteína Cas9 catalíticamente inactiva (proteína dCas9) que no escinde el ácido nucleico alejado del objetivo, y (ii) un segundo polinucleótido guía que comprende un espaciador adaptado para unirse al ácido nucleico alejado del objetivo;

lo que reduce la unión y/o la escisión por el primer complejo del ácido nucleico alejado del objetivo, en el que el ácido nucleico objetivo seleccionado es ADN, y

en el que el procedimiento no comprende el tratamiento del cuerpo humano o animal mediante cirugía o terapia practicada en el cuerpo humano o animal.
- 15 2. El procedimiento de la reivindicación 1, en el que el primer polinucleótido guía es un ARN guía único y el segundo polinucleótido guía es un ARN guía único.
- 20 3. El procedimiento de la reivindicación 1 o la reivindicación 2, en el que la proteína Cas9 catalíticamente activa es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga.
- 25 4. El procedimiento de una cualquiera de las reivindicaciones 1 a 3, en el que la proteína dCas9 comprende al menos una mutación en uno o más dominios de endonucleasa para volver la proteína dCas9 catalíticamente inactiva.
5. El procedimiento de la reivindicación 4, en el que la proteína dCas9 es una proteína Cas9 de *S. pyogenes* o una proteína Cas9 ortóloga, y la proteína dCas9 comprende al menos una mutación en uno o más dominios de endonucleasa para volver la proteína Cas9 ortóloga catalíticamente inactiva.
6. El procedimiento de la reivindicación 4 o la reivindicación 5, en el que la al menos una mutación está en un dominio RuvC-1.
7. El procedimiento de una cualquiera de las reivindicaciones 4, 5 o 6, en el que la al menos una mutación está en un dominio HNH.
- 30 8. El procedimiento de cualquier reivindicación precedente, en el que el ácido nucleico objetivo seleccionado se escinde para proporcionar un sitio de escisión, y el procedimiento comprende además la modificación del ácido nucleico objetivo.
9. El procedimiento de la reivindicación 8, en el que la modificación comprende insertar al menos una parte de un polinucleótido donante en el sitio de escisión.
- 35 10. El procedimiento de la reivindicación 8, en el que la modificación comprende eliminar uno o más nucleótidos en el sitio de escisión.
11. El procedimiento de cualquier reivindicación precedente, en el que el procedimiento se realiza en una célula.
12. El procedimiento de la reivindicación 11, en el que la célula es una célula eucariota.
13. El procedimiento de una cualquiera de las reivindicaciones 1 a 10, en el que el procedimiento se realiza in vitro.
- 40 14. El procedimiento de cualquier reivindicación precedente, en el que el ácido nucleico objetivo seleccionado es ADN de cadena doble.



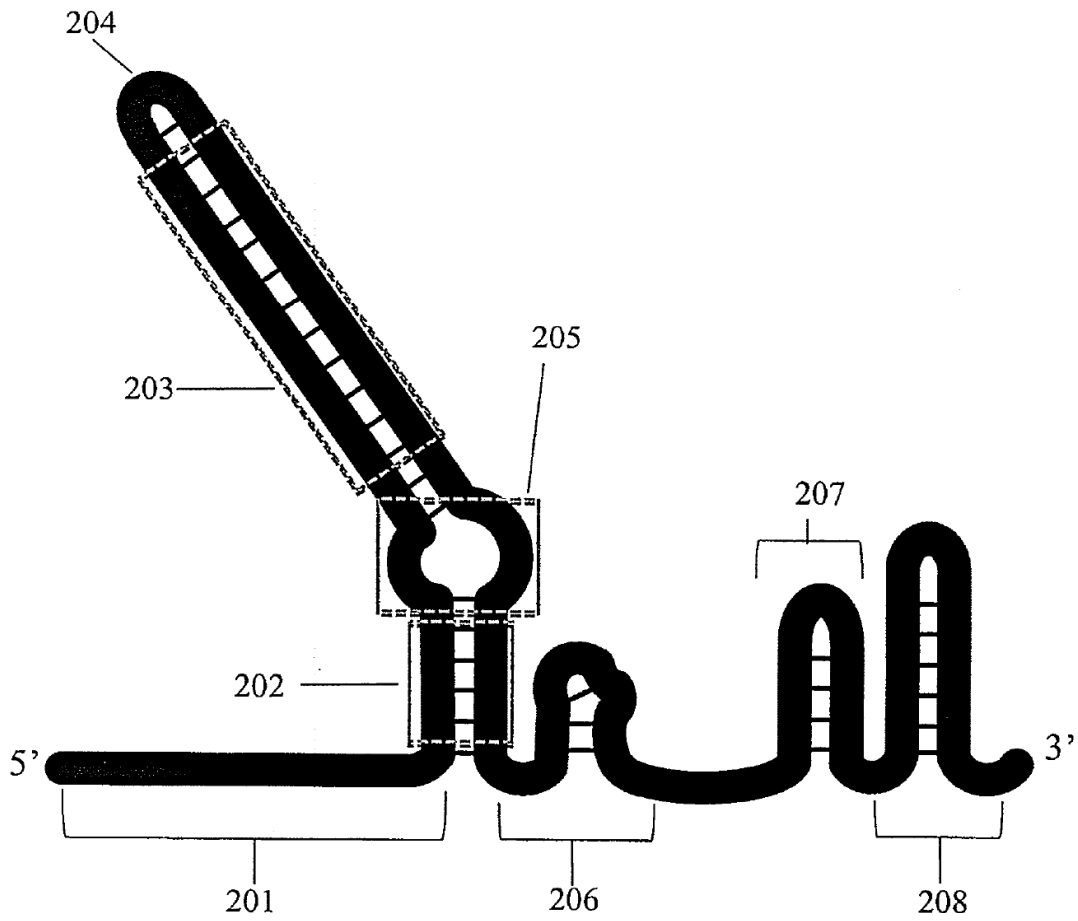


FIG. 2

FIG. 3A

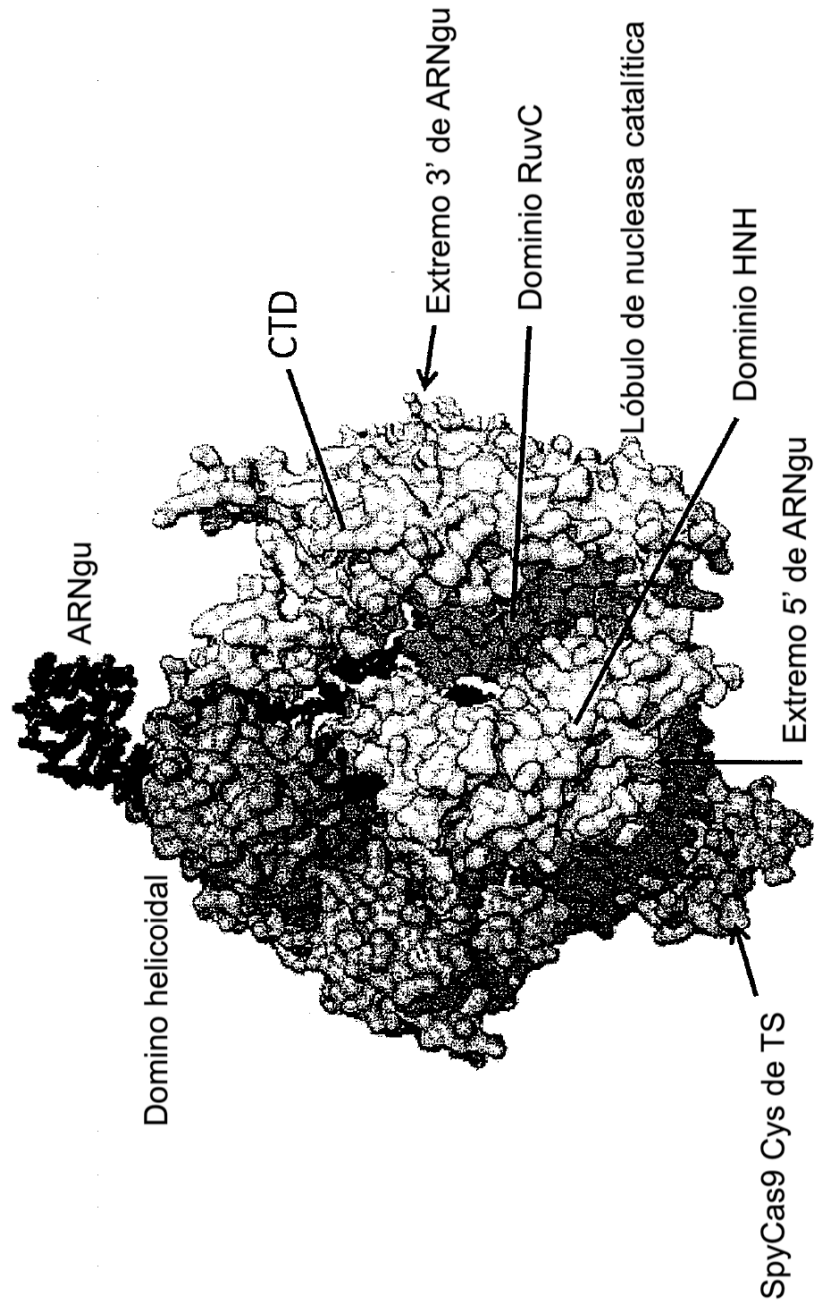


FIG. 3B

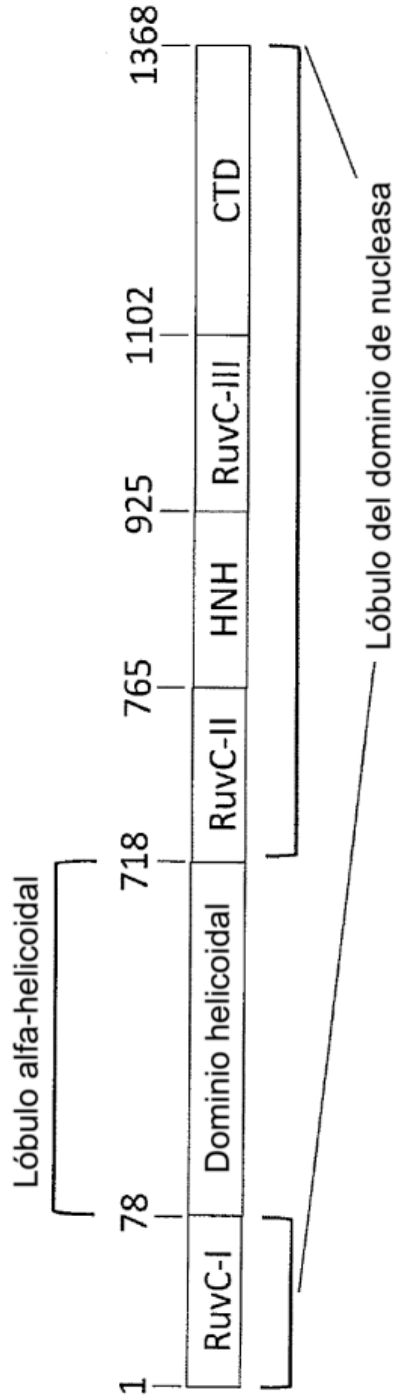


FIG. 4

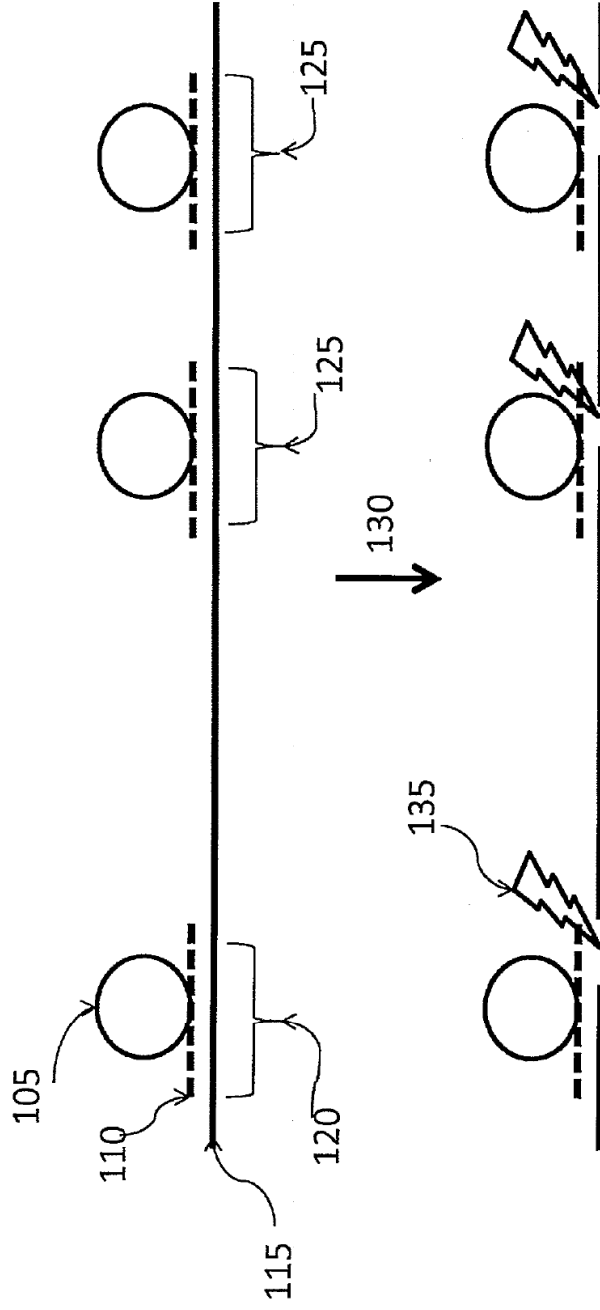
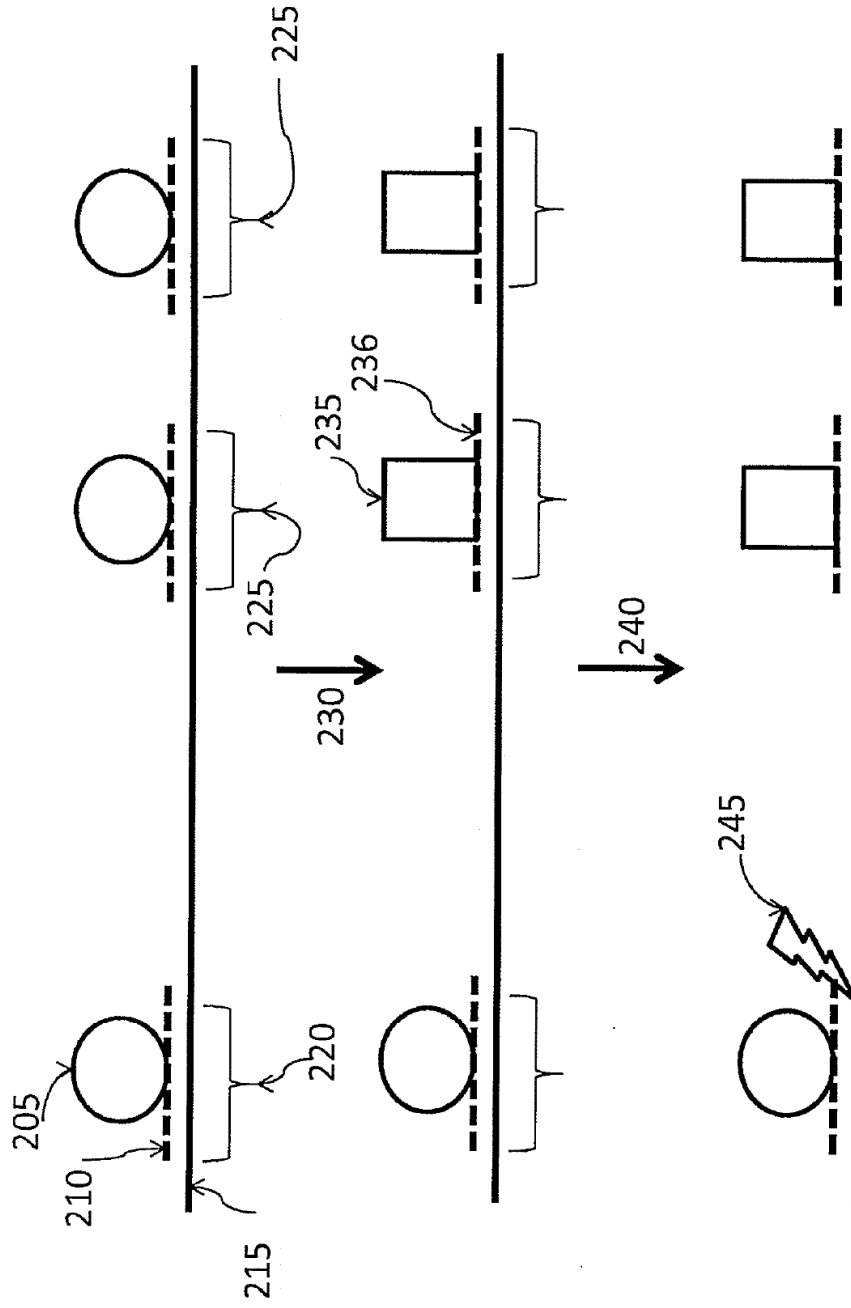


FIG. 5



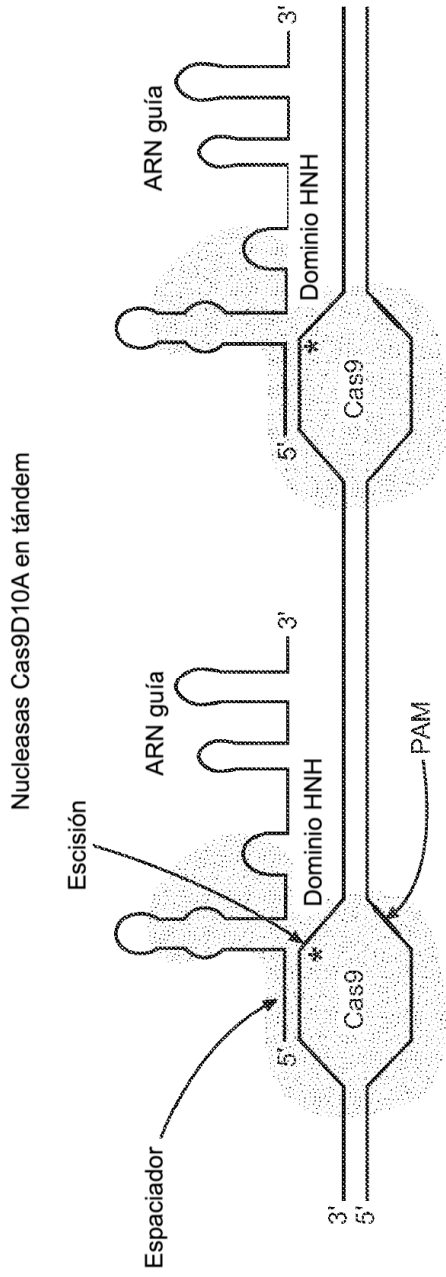


FIG. 6A



FIG. 6B

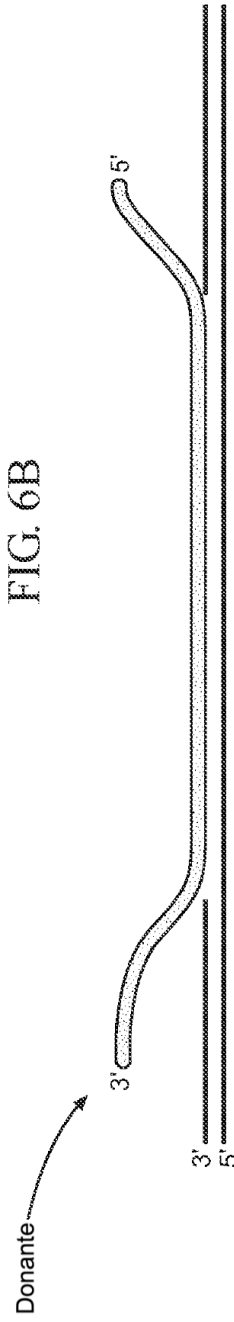


FIG. 6C

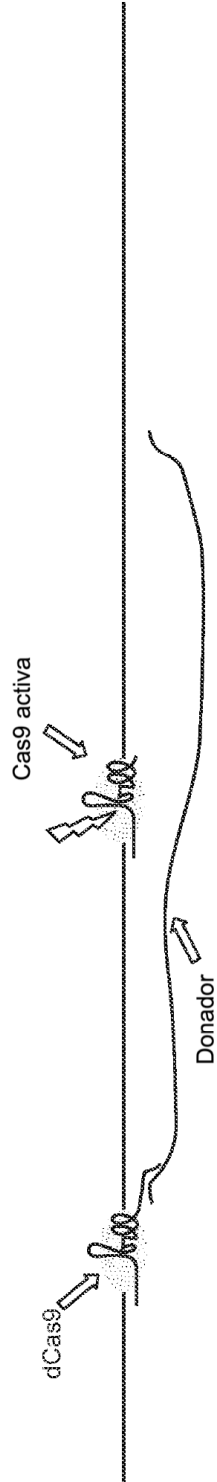


FIG. 7A

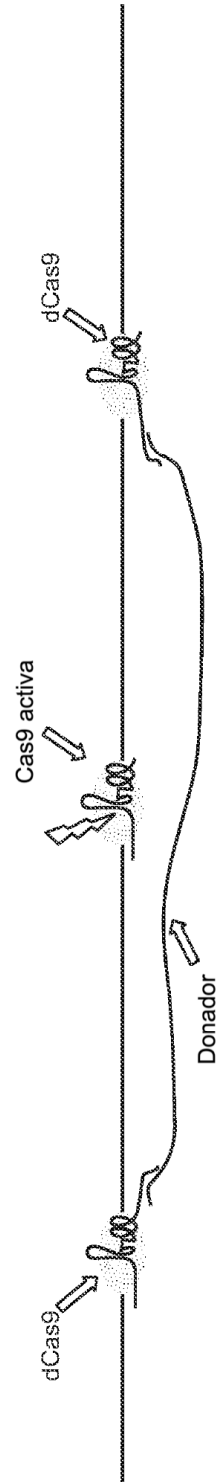


FIG. 7B

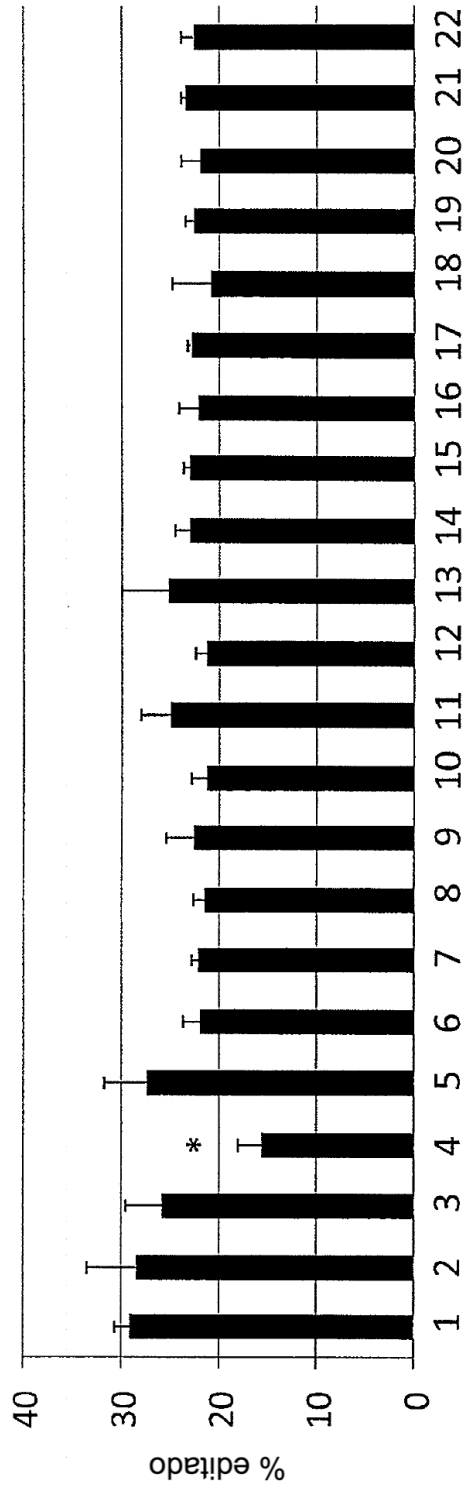


FIG. 8

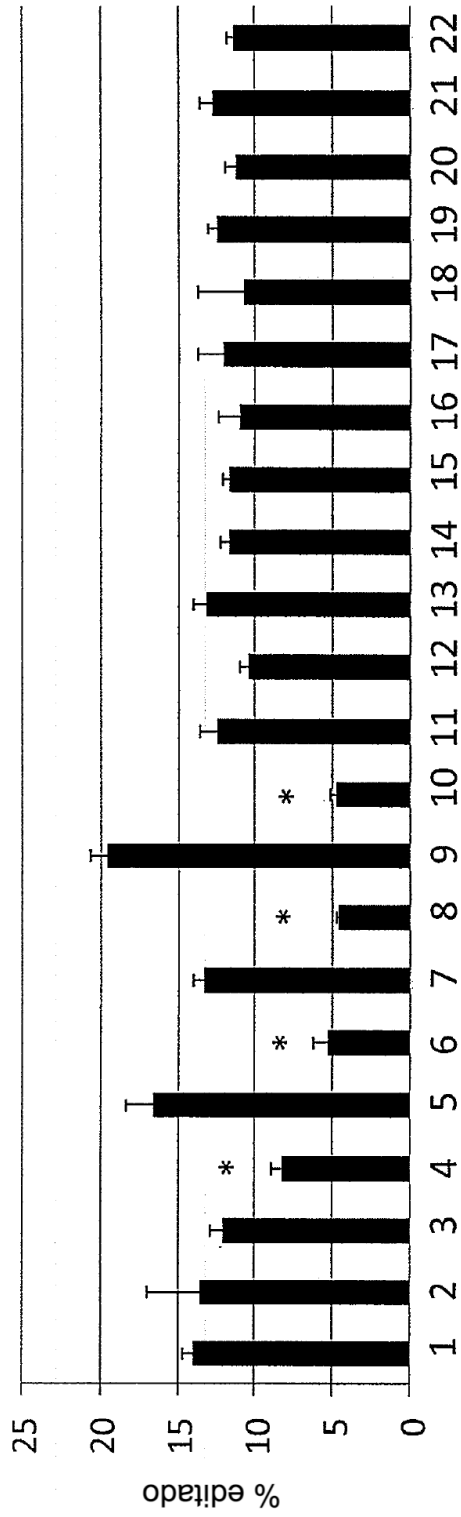


FIG. 9

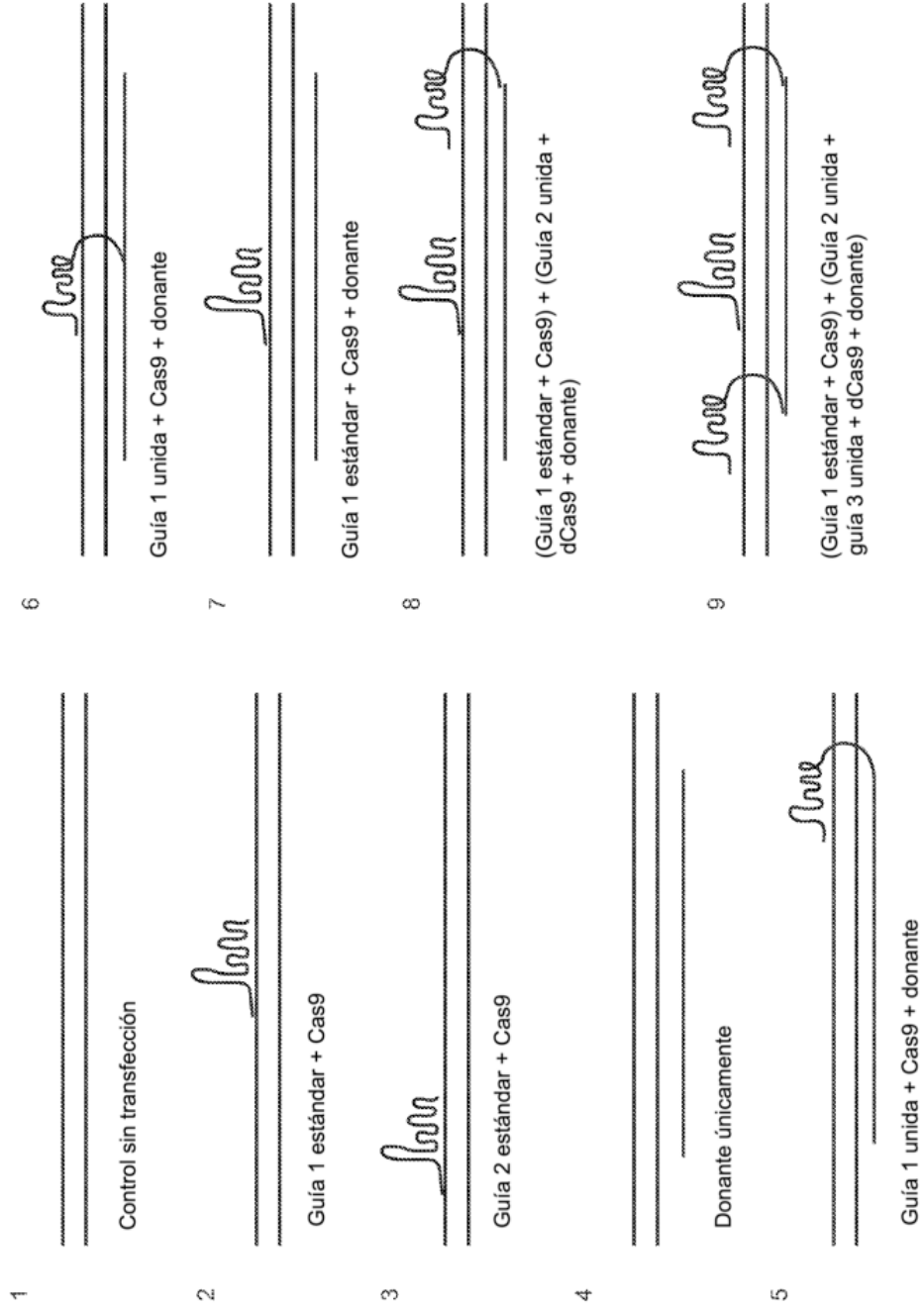


FIG. 10

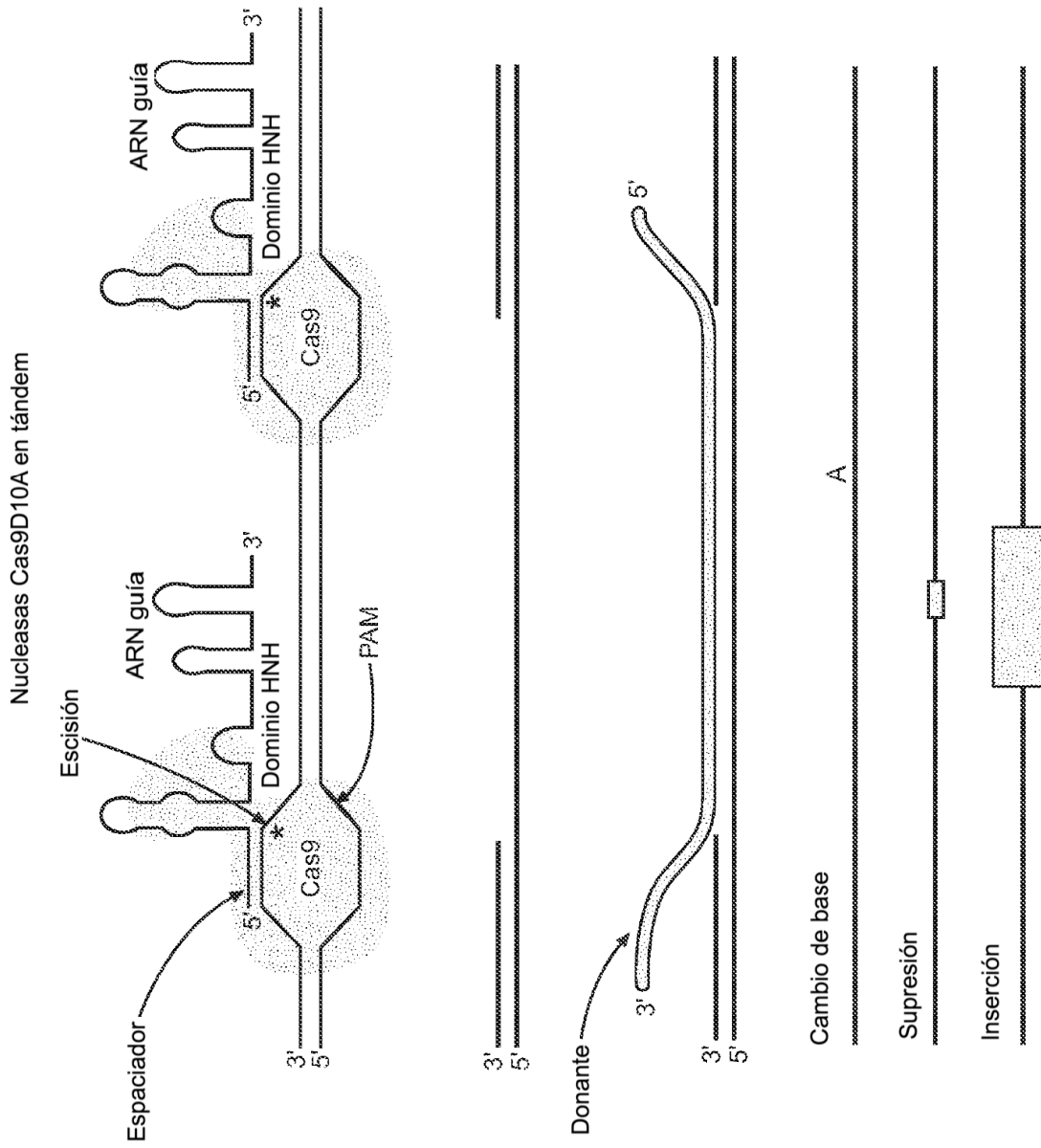


FIG. 11

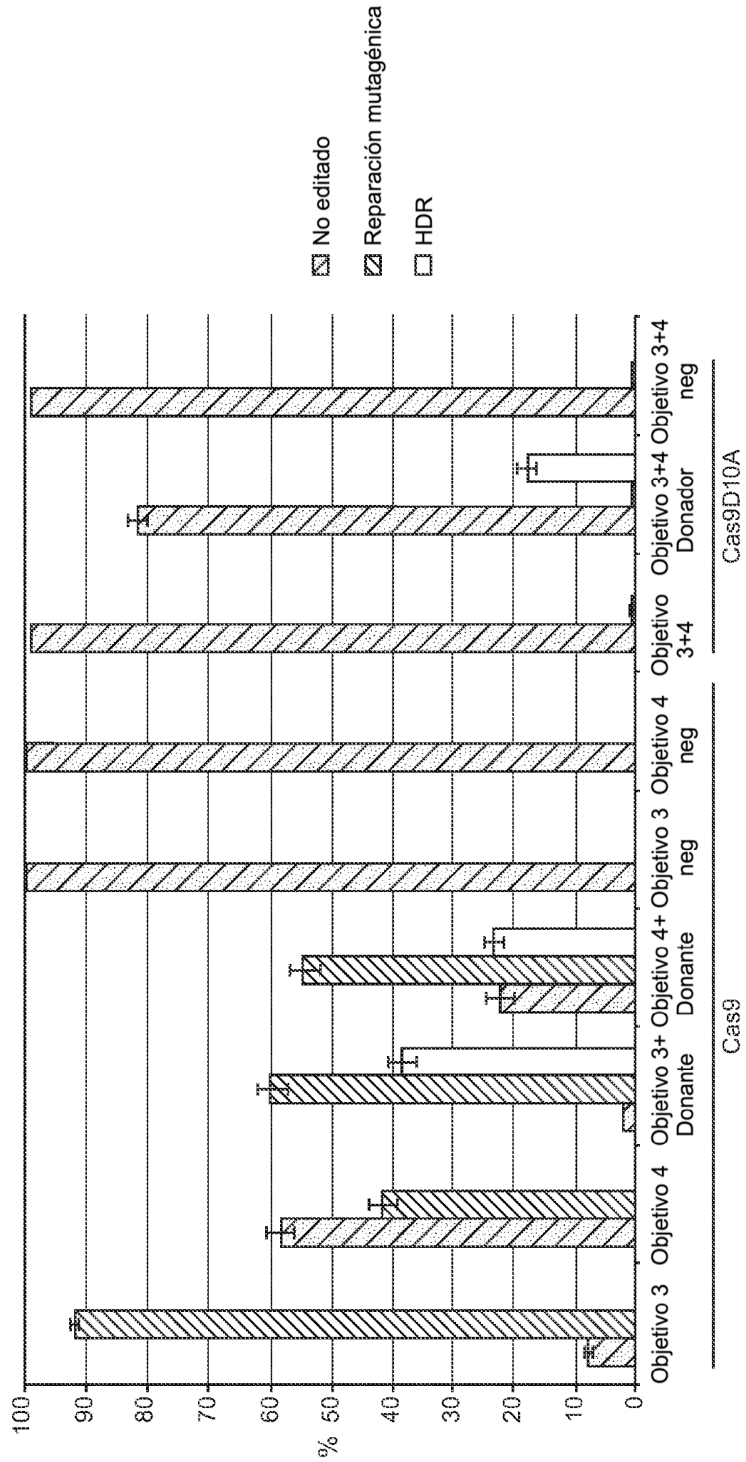


FIG. 12