

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 760 477**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 9/22** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.02.2017 E 17157024 (5)**

97 Fecha y número de publicación de la concesión europea: **25.09.2019 EP 3272867**

54 Título: **Uso de proteínas de unión a ADN programables para potenciar la modificación dirigida del genoma**

30 Prioridad:

**02.06.2016 US 201662344858 P**

**05.07.2016 US 201662358415 P**

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

**14.05.2020**

73 Titular/es:

**SIGMA ALDRICH CO. LLC (100.0%)**  
**3050 Spruce Street**  
**St. Louis, MO 63103, US**

72 Inventor/es:

**CHEN, FUQIANG**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

ES 2 760 477 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

Uso de proteínas de unión a ADN programables para potenciar la modificación dirigida del genoma

**Campo**

5 La presente divulgación se refiere a composiciones y procedimientos para aumentar la eficacia y/o la especificidad de la modificación dirigida del genoma.

**Antecedentes**

10 Las endonucleasas programables se han convertido en una herramienta importante para la ingeniería o modificación dirigida del genoma en eucariotas. Recientemente, los sistemas de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR, forma siglada de *clustered regularly interspersed short palindromic repeats*)/asociados a CRISPR (Cas, forma siglada de *CRISPR-associated*) (CRISPR/Cas) guiados por ARN han surgido como una nueva generación de herramientas para la modificación del genoma. Estas nuevas endonucleasas programables han mejorado enormemente la capacidad de edición del genoma en comparación con las generaciones anteriores de nucleasas, tales como las nucleasas de dedos de zinc (las ZFN, forma siglada de *zinc finger nuclease*) y las nucleasas efectoras de tipo activador de la transcripción (las TALEN, forma siglada *transcription activator-like effector nucleases*).

15 Sin embargo, no todas las dianas genómicas son accesibles para una modificación eficaz por estas endonucleasas programables. De hecho, algunas endonucleasas CRISPR-Cas parecen tener poca o ninguna actividad en las células humanas. Entre otras cosas, la estructura de la cromatina puede presentar una barrera para estas endonucleasas programables y evitar que se unan a la secuencia diana. Por lo tanto, existe la necesidad de mejorar la accesibilidad de estas endonucleasas programables a las secuencias diana y/o de mejorar la eficacia de la modificación dirigida del genoma. Además, existe la necesidad de aumentar la especificidad de la modificación dirigida del genoma reduciendo los efectos inespecíficos.

20 Richardson y col. (Nature Biotech., 2016, 34(3):339-344) desvelan procedimientos para potenciar la edición del genoma dirigida por homología mediante CRISPR-Cas9 catalíticamente activa e inactiva, utilizando ADN donante asimétrico. El documento WO2016/036754 desvela, entre otros, composiciones y procedimientos para la modificación específica de un ADN diana en una célula eucariota. Ran y col. (Cell, 2013, 154(6):1380-1389) desvelan el doble mellado de ácido nucleico diana mediante construcciones de CRISPR-Cas9, para mejorar la especificidad de edición del genoma. Los documentos WO2014/191518 y WO2016073990 también desvelan construcciones y procedimientos de doble mellado similares. Los documentos WO2014/089290 y WO2016/007604 desvelan la fusión de la endonucleasa FokI con una nucleasa Cas9 inactiva y el uso de la fusión para escindir el ADN. Los documentos  
30 WO2017/070598 y WO2017/096328 desvelan construcciones específicas de ácido nucleico dirigido a ácido nucleico (NATNA, forma siglada de *nucleic-acid targeting nucleic acid*), que unen varios componentes del sistema para reducir el número de partes discretas, con la intención de reducir el número de componentes en el sistema.

**Sumario**

Entre los diversos aspectos de la presente divulgación se encuentra una composición que comprende:

35 (a) un sistema de nucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR, forma siglada de *clustered regularly interspersed short palindromic repeats*) guiado por ARN o un ácido nucleico que codifica dicho sistema de nucleasa CRISPR, en el que el sistema de nucleasa CRISPR comprende (i) una proteína de modificación de ADN programable que es una proteína CRISPR, y (ii) un ARN de guía; y

40 (b) al menos un sistema CRISPR catalíticamente inactivo o un ácido nucleico que codifica dicho al menos un sistema CRISPR catalíticamente inactivo, en el que cada sistema CRISPR catalíticamente inactivo comprende (i) una proteína de unión a ADN programable que es una proteína CRISPR catalíticamente inactiva y (ii) un ARN de guía; y en la que

(x) la proteína de modificación de ADN programable es una proteína CRISPR de tipo II y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II, o

45 (y) la proteína de modificación de ADN programable es una proteína CRISPR de tipo V y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II; y en la que la composición como se define anteriormente no se desvela en los documentos WO2017/070598 o WO2017/096328.

50 En general, el ácido nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable es ARNm o ADN. En algunas realizaciones, el ácido nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable es parte de un vector tal como, por ejemplo, un vector plasmídico, un vector lentivírico, un vector vírico adenoasociado o un vector adenovírico.

En realizaciones específicas, la proteína de modificación de ADN programable comprende un sistema de nucleasa CRISPR/Cas, o un sistema de nickasa doble CRISPR/Cas, y la al menos una proteína de unión a ADN programable

comprende un sistema CRISPR/Cas catalíticamente inactivo, en el que cada sistema CRISPR/Cas comprende una proteína CRISPR/Cas y un ARN de guía. En algunas realizaciones, cada ARN de guía puede sintetizarse químicamente al menos de forma parcial. En otras realizaciones, cada ARN de guía puede sintetizarse enzimáticamente. En realizaciones adicionales, el ácido nucleico que codifica cada proteína CRISPR/Cas puede ser ARNm, y el ácido nucleico que codifica cada ARN de guía puede ser ADN. En aún otras realizaciones, el ácido nucleico que codifica cada proteína CRISPR/Cas puede ser ARNm, y el ácido nucleico que codifica cada ARN de guía puede ser ADN. En determinados aspectos, el ácido nucleico que codifica la proteína CRISPR/Cas y/o el ácido nucleico que codifica el ARN de guía puede ser parte de un vector, por ejemplo, un vector plasmídico, un vector lentivírico, un vector vírico adenoasociado o un vector adenovírico.

5 Otro aspecto de la presente divulgación abarca kits que comprenden una cualquiera o más de las composiciones detalladas anteriormente.

Aún otro aspecto de la presente divulgación proporciona un procedimiento *in vitro* para aumentar la eficacia y/o especificidad de la modificación dirigida del genoma en una célula eucariota, comprendiendo el procedimiento introducir en la célula eucariota una composición que comprende:

15 (a) un sistema de nucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) guiado por ARN o un ácido nucleico que codifica dicho sistema de nucleasa CRISPR, en el que el sistema de nucleasa CRISPR comprende (i) una proteína de modificación de ADN programable que es una proteína CRISPR, y (ii) un ARN de guía; y

20 (b) al menos un sistema CRISPR catalíticamente inactivo o un ácido nucleico que codifica dicho al menos un sistema CRISPR catalíticamente inactivo, en el que cada sistema CRISPR catalíticamente inactivo comprende (i) una proteína de unión a ADN programable que es una proteína CRISPR catalíticamente inactiva y (ii) un ARN de guía; y en la que

(x) la proteína de modificación de ADN programable es una proteína CRISPR de tipo II y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II, o

25 (y) la proteína de modificación de ADN programable es una proteína CRISPR de tipo V y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II; en la que la proteína de modificación de ADN programable se dirige a una secuencia cromosómica diana y cada una de las al menos una proteína de unión a ADN programable se dirige a un sitio próximo a la secuencia cromosómica diana, y la unión de la al menos una proteína de unión a ADN programable al sitio próximo a la secuencia cromosómica diana aumenta la accesibilidad de la proteína de modificación de ADN programable a la secuencia cromosómica diana, aumentando de este modo la eficacia y/o la especificidad de la modificación dirigida del genoma.

30 Se localiza el sitio próximo al que se une cada una de las al menos una proteína de unión a ADN programable, por ejemplo, dentro de aproximadamente 250 pares de bases a cada lado de la secuencia cromosómica diana. En algunas realizaciones, el sitio (o sitios) de unión próximos están ubicados a menos de aproximadamente 200 pb o menos de aproximadamente 100 pb a cada lado de la secuencia cromosómica diana.

35 En realizaciones específicas, la proteína de modificación de ADN programable comprende un sistema de nucleasa CRISPR/Cas, o un sistema de nickasa doble CRISPR/Cas, y la al menos una proteína de unión a ADN programable comprende un sistema CRISPR/Cas catalíticamente inactivo, en el que cada sistema CRISPR/Cas comprende una proteína CRISPR/Cas y un ARN de guía.

40 En diversas realizaciones, se introducen en la célula eucariota al menos dos, al menos tres, o más de tres proteínas de unión a ADN programables. En realizaciones específicas, la célula eucariota es una célula de mamífero o una célula humana.

Otros aspectos y características de la divulgación se detallan a continuación.

### **Breve descripción de los dibujos**

45 La FIG. 1 proporciona un diagrama de una realización de los procedimientos desvelados en el presente documento. La unión próxima de una proteína (o proteínas) de unión a ADN programable aumenta la accesibilidad al sitio diana para una nucleasa programable, aumentando de este modo la eficacia de la escisión en el sitio diana.

50 La FIG. 2 ilustra que la unión de SpCas9 catalíticamente inactiva (SpdCas9) a un sitio (o sitios) próximo aumenta la eficacia de escisión mediante FnCas9. Las secuencias presentadas en la parte superior muestran las ubicaciones relativas del sitio diana de FnCas9 en el locus POR y los sitios de unión de SpdCas9. En la parte inferior se muestran los resultados de un ensayo de nucleasa Cel-I.

La FIG. 3A ilustra el diseño de un experimento para determinar si la unión de SpCas9 catalíticamente inactiva (SpdCas9) aumenta la accesibilidad y la unión de CjCas9 catalíticamente inactiva (CjdCas9) etiquetada con epítipo (es decir, etiquetada con FLAG®) a un sitio previamente inaccesible en el locus POR.

55 La FIG. 3B proporciona un diagrama del ensayo de unión de inmunoprecipitación de cromatina utilizado para detectar la unión de CjdCas9 etiquetada con epítipo a sitios diana en los loci POR y AAVS1.

La FIG. 3C ilustra que la unión de SpdCas9 a sitios próximos aumenta la unión de CjCas9 etiquetada con epítipo a un sitio previamente inaccesible en el locus POR.

La FIG. 4 ilustra que la unión de SpCas9 catalíticamente inactiva (SpdCas9) a un sitio (o sitios) próximo aumenta la eficacia de escisión mediante CjCas9. Las secuencias presentadas en la parte superior muestran las ubicaciones relativas del sitio diana de CjCas9 en el locus POR y los sitios de unión de SpdCas9. En la parte inferior se muestran los resultados de un ensayo de nucleasa Cel-I.

La FIG. 5 ilustra que la unión de SpCas9 catalíticamente inactiva (SpdCas9) a un sitio (o sitios) próximo aumenta la eficacia de escisión mediante FnCpf1. Las ubicaciones relativas del sitio diana de FnCpf1 y los sitios de unión de SpdCas9 en el locus POR se ilustran en la parte superior y los resultados de un ensayo de nucleasa Cel-I se muestran en la parte inferior.

La FIG. 6 ilustra que la unión de SpCas9 catalíticamente inactiva (SpdCas9) a un sitio (o sitios) próximo aumenta la escisión específica mediante CjCas9. Los sitios diana de CjCas9 en los locus HBD y HBB, así como los sitios de unión de SpdCas9 en el locus HBB, se muestran en la parte superior. En la parte inferior se muestran los resultados de un ensayo de nucleasa Cel-I.

La FIG. 7 ilustra que la unión de FnCas9 catalíticamente inactiva (FndCas9) a un sitio (o sitios) próximo aumenta la escisión específica mediante SpCas9. Las ubicaciones relativas del sitio diana de SpCas9 y los sitios de unión de FndCas9 en el locus POR se indican en la parte superior. En la parte inferior se muestran los resultados de un ensayo de nucleasa Cel-I.

La FIG. 8 ilustra la potenciación de la edición de genes mediada por oligonucleótidos de ADNmc. Las ubicaciones relativas de los sitios diana en el locus POR y la secuencia del oligonucleótido de ADNmc se muestran en la parte superior. En la parte inferior se muestran los resultados de la integración dirigida del sitio EcoRI. Las eficacias de integración (%) del sitio EcoRI se determinaron por ImageJ. M: Marcadores de ADN de amplia gama. ND: no determinado.

### **Descripción detallada**

La presente divulgación proporciona composiciones y procedimientos para aumentar la accesibilidad del ADN cromosómico a endonucleasas dirigidas y otras proteínas de modificación de ADN programables, en los que la accesibilidad aumentada conduce a una eficacia y/o especificidad aumentadas de la modificación del genoma o la modificación epigenética dirigida. Se ha descubierto que algunas endonucleasas CRISPR/Cas tienen actividad reducida o nula en células humanas. Es posible que la ocupación de nucleosomas, el posicionamiento y la forma en que una secuencia de ADN se envuelve alrededor del octámero de histona pueda determinar qué tan accesible es la secuencia para una proteína de unión a ADN (Chereji y col., Briefing Functional Genomics, 2014, 14:506-60). Por lo tanto, es posible que el impedimento impuesto por la configuración local de la cromatina pueda desempeñar un papel en la aparente inactividad de muchas endonucleasas CRISPR/Cas en células humanas. Se ha descubierto, como se detalla en el presente documento, que la unión de las proteínas de unión a ADN a sitios ubicados próximos (es decir, dentro de aproximadamente 250 pares de bases) al sitio diana de una proteína de modificación de ADN dirigida aumenta la accesibilidad de la proteína de modificación de ADN dirigida al sitio diana, aumentando de este modo la eficacia y/o especificidad de la modificación del genoma dirigida o la modificación epigenética dirigida. Las composiciones y los procedimientos desvelados en el presente documento, por lo tanto, permite la modificación del genoma/modificación epigenética dirigida eficaz utilizando endonucleasas CRISPR/Cas que anteriormente se pensaba que eran inactivas en células humanas. Además, las composiciones y los procedimientos desvelados en el presente documento también mejoran la modificación selectiva del genoma entre sitios diana casi idénticos, reduciendo de este modo los efectos inespecíficos.

### ***(I) Composiciones***

Un aspecto de la presente divulgación proporciona composiciones que comprenden (a) proteínas de modificación de ADN programables o un ácido nucleico que codifica las proteínas de modificación de ADN programables y (b) al menos una proteína de unión a ADN programable o un ácido nucleico que codifica la al menos una proteína de unión a ADN programable. Las proteínas de modificación de ADN programables se detallan a continuación en la sección (I)(a), las proteínas de unión a ADN programables se detallan a continuación en la sección (I)(b), y los ácidos nucleicos que codifican estas proteínas se detallan a continuación en la sección (I)(c).

### **(a) Proteínas de modificación de ADN programables**

Una proteína de modificación de ADN programable es una proteína que se une a una secuencia diana específica en el ADN cromosómico y modifica el ADN o una proteína asociada con el ADN en o cerca de la secuencia diana. Por lo tanto, una proteína de modificación de ADN programable comprende un dominio de unión a ADN y un dominio de modificación catalíticamente activo.

El dominio de unión a ADN es programable, en el sentido que puede diseñarse o genomanipularse para reconocer y unirse a distintas secuencias de ADN. En algunas realizaciones, por ejemplo, la unión al ADN está mediada por la interacción entre la proteína y el ADN diana. Por lo tanto, el dominio de unión a ADN puede programarse para unirse a una secuencia de ADN de interés mediante ingeniería de proteínas. En otras realizaciones, por ejemplo, la unión a ADN está mediada por un ARN de guía que interactúa con el dominio de unión a ADN programable de la proteína y el ADN diana. En determinados casos, el dominio de unión a ADN programable puede dirigirse a una secuencia de ADN de interés diseñando el ARN de guía apropiado.

Se puede incluir una diversidad de dominios de modificación en las proteínas de modificación de ADN programables. En algunas realizaciones, el dominio de modificación es un dominio nucleasa, que tiene actividad nucleasa y escinde ambas cadenas de una secuencia de ADN bicatenario (es decir, genera una rotura bicatenaria). La rotura bicatenaria puede repararse luego mediante un proceso celular de reparación de ADN, tal como por unión de extremos no homólogos (NHEJ, forma siglada de *non-homologous end-joining*) o la reparación dirigida por homología (HDR, forma siglada de *homology-directed repair*). Como consecuencia, la secuencia de ADN puede modificarse mediante una delección, inserción y/o sustitución de al menos un par de bases hasta, por ejemplo, muchos miles de pares de bases. Los ejemplos de proteínas de modificación de ADN programables que comprenden dominios nucleasa incluyen, sin limitación, sistemas de nucleasa CRISPR/Cas, sistemas de nickasa doble CRISPR/Cas, nucleasas de dedos de zinc, nucleasas efectoras de tipo activador de la transcripción, meganucleasas, proteínas de fusión que comprenden un dominio de nucleasa unido a un dominio de unión a ADN programable, y combinaciones de los mismos. Las proteínas de modificación de ADN programables que comprenden dominios nucleasa se detallan a continuación en las secciones (I)(a)(i)-(ii).

Las proteínas de modificación de ADN programables pueden comprender dominios de unión y/o modificación de ADN de tipo origen natural o de tipo silvestre, versiones modificadas de dominios de unión y/o modificación de ADN de origen natural, dominios de unión y/o modificación de ADN sintéticos o artificiales, y combinaciones de los mismos.

#### (i) Sistemas de nucleasa CRISPR/Cas

En algunas realizaciones, la proteína de modificación de ADN programable puede ser un sistema de nucleasa CRISPR/Cas guiado por ARN, que introduce una rotura bicatenaria en el ADN. El sistema de nucleasa CRISPR/Cas comprende una nucleasa de CRISPR/Cas y un ARN de guía.

Nucleasa CRISPR/Cas. La nucleasa CRISPR/Cas puede obtenerse de un sistema CRISPR de tipo I (es decir, IA, IB, IC, ID, IE o IF), tipo II (es decir, IIA, IIB o IIC), tipo III (es decir, IIIA o IIIB), o tipo V, que están presentes en diversas bacterias y arqueas. Por ejemplo, el sistema CRISPR/Cas puede ser de *Streptococcus* sp. (por ejemplo, *Streptococcus pyogenes*), *Campylobacter* sp. (por ejemplo, *Campylobacter jejuni*), *Francisella* sp. (por ejemplo, *Francisella novicida*), *Acaryochloris* sp., *Acetohalobium* sp., *Acidaminococcus* sp., *Acidithiobacillus* sp., *Alicyclobacillus* sp., *Allochromatium* sp., *Ammonifex* sp., *Anabaena* sp., *Arthrospira* sp., *Bacillus* sp., *Burkholderiales* sp., *Caldicelulosiruptor* sp., *Candidatus* sp., *Clostridium* sp., *Crocospaera* sp., *Cyanothece* sp., *Exiguobacterium* sp., *Finegoldia* sp., *Ktedonobacter* sp., *Lachnospiraceae* sp., *Lactobacillus* sp., *Lyngbya* sp., *Marinobacter* sp., *Methanohalobium* sp., *Microscilla* sp., *Microcoleus* sp., *Microcystis* sp., *Natranaerobius* sp., *Neisseria* sp., *Nitrosococcus* sp., *Nocardioopsis* sp., *Nodularia* sp., *Nostoc* sp., *Oscillatoria* sp., *Polaromonas* sp., *Pelotomaculum* sp., *Pseudoalteromonas* sp., *Petrogona* sp., *Prevotella* sp., *Staphylococcus* sp., *Streptomyces* sp., *Streptosporangium* sp., *Synechococcus* sp., *Thermosiphon* sp., o *Verrucomicrobia* sp. En aún otras realizaciones, la nucleasa CRISPR/Cas puede obtenerse de un sistema CRISPR de arqueas, un sistema CRISPR-CasX o un sistema CRISPR-CasY (Burststein y col., Nature, 2017, 542(7640):237-241).

En una realización particular, la nucleasa CRISPR/Cas puede obtenerse de un sistema CRISPR/Cas de tipo II. En otra realización particular, la nucleasa CRISPR/Cas puede obtenerse de un sistema CRISPR/Cas de tipo V.

Los ejemplos no limitantes de proteínas CRISPR adecuadas incluyen proteínas Cas, proteínas Cpf, proteínas C2c (por ejemplo, C2c1, C2c2, Cdc3), proteínas Cmr, proteínas Csa, proteínas Csb, proteínas Csc, proteínas Cse, proteínas Csf, proteínas Csm, proteínas Csn, proteínas Csx, proteínas Csy, proteínas Csz, y derivados o variantes de las mismas. En realizaciones específicas, la nucleasa CRISPR/Cas puede ser una proteína Cas9 de tipo II, una proteína Cpf1 de tipo V, o un derivado de las mismas.

En algunas realizaciones, la nucleasa CRISPR/Cas puede ser Cas9 de *Streptococcus pyogenes* (SpCas9) o Cas9 de *Streptococcus thermophilus* (StCas9). En otras realizaciones, la nucleasa CRISPR/Cas puede ser una Cas9 de *Campylobacter jejuni* (CjCas9). En realizaciones alternativas, la nucleasa CRISPR/Cas puede ser una Cas9 de *Francisella novicida* (FnCas9). En aún otras realizaciones, la nucleasa CRISPR/Cas puede ser una Cas9 de *Neisseria cinerea* (NcCas9). En realizaciones adicionales, la nucleasa CRISPR/Cas puede ser una Cpf1 de *Francisella novicida* (FnCpf1), Cpf1 de *Acidaminococcus* sp. (AsCpf1) o Cpf1 de *Lachnospiraceae bacterium* ND2006 (LbCpf1).

En general, la nucleasa CRISPR/Cas comprende un dominio de reconocimiento de ARN y/o de unión a ARN, que interactúa con el ARN de guía. La nucleasa CRISPR/Cas también comprende al menos un dominio de nucleasa que tiene actividad endonucleasa. Por ejemplo, una proteína Cas9 comprende un dominio de nucleasa de tipo RuvC y un dominio de nucleasa de tipo HNH, y una proteína Cpf1 comprende un dominio de tipo RuvC. Las nucleasas CRISPR/Cas también pueden comprender dominios de unión a ADN, dominios de helicasa, dominios de ARNasa, dominios de interacción proteína-proteína, dominios de dimerización, así como otros dominios.

La nucleasa CRISPR/Cas puede comprender, adicionalmente, al menos una señal de localización nuclear, un dominio de penetración celular y/o un dominio marcador. Los ejemplos no limitantes de señales de localización nuclear incluyen PKKKRKV (SEQ ID NO: 1), PKKKRRV (SEQ ID NO:2), KRPAATKKAGQAKKKK (SEQ ID NO:3), YGRKKRRQRRR (SEQ ID NO:28), RKKRRQRRR (SEQ ID NO:29), PAAKRVKLD (SEQ ID NO:30), RQRRNELKRSP (SEQ ID NO:31), VSRKRPRP (SEQ ID NO:32), PPKKARED (SEQ ID NO:33), PPKKKPL (SEQ ID NO:34), SALIKKKKKMAP (SEQ ID

NO:35), PKQKKRK (SEQ ID NO:36), RKLKKIKKL (SEQ ID NO:37), REKKKFLKRR (SEQ ID NO:38), KRKGDEVDGVDEVAKKSKK (SEQ ID NO:39), RKCLQAGMNLEARKTKK (SEQ ID NO:40), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:41) y RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:42). Los ejemplos de dominios de penetración celular adecuados incluyen, sin limitación, GRKKRRQRRRPPQPKKKRKV (SEQ ID NO:4), PLSSIFSRIGDPPKKRKV (SEQ ID NO:5), GALFLGWLGAAGSTMGAPKKRKV (SEQ ID NO:6), GALFLGFLGAAGSTMGAWSQPKKRKV (SEQ ID NO: 7), KETWWETWWTEWSQPKKRKV (SEQ ID NO: 8), YARAAARQARA (SEQ ID NO:43), THRLPRRRRRR (SEQ ID NO:44), GGRRARRRRR (SEQ ID NO:45), RRQRRTSKLMKR (SEQ ID NO:46), GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:47), KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:48) y RQIKIWFQNRRMKWKK (SEQ ID NO:49). Los dominios marcadores incluyen proteínas fluorescentes y etiquetas de purificación o de epítipo. Las proteínas fluorescentes adecuadas incluyen, sin limitación, proteínas verdes fluorescentes (por ejemplo, GFP, eGFP, GFP-2, tagGFP, turboGFP, Emerald, Verde Azami, Verde Azami monomérico, CopGFP, AceGFP, ZsGreen1), proteínas amarillas fluorescentes (por ejemplo, YFP, EYFP, Citrina, Venus, YPet, PhiYFP, ZsYellow1), proteínas azules fluorescentes (por ejemplo, BFP, EBFP, EBFP2, Azurite, mKama1, GFPuv, Sapphire, T-sapphire), proteínas cian fluorescentes (por ejemplo, ECFP, Cerúlea, CyPet, AmCyan1, Midoriishi-Cyan), proteínas rojas fluorescentes (por ejemplo, mKate, mKate2, mPlum, monómero de DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, monómero de DsRed, Tándem de HcRed, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred) y proteínas naranjas fluorescentes (por ejemplo, mOrange, mKO, Kusabira-Orange, Kusabira-Orange monomérica, mTangerine, tdTomato). Los ejemplos no limitantes de etiquetas de purificación o de epítipo adecuadas incluyen 6xHis, FLAG®, HA, GST, Myc, y similares.

La señal de localización nuclear, el dominio de penetración celular y/o el dominio marcador pueden ubicarse en el extremo N, el C-terminal, o en una ubicación interna de la proteína. En algunas realizaciones, la nucleasa CRISPR/Cas puede comprender adicionalmente al menos un marcador detectable. El marcador detectable puede ser un fluoróforo (por ejemplo, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluor, etiquetas Halo o una etiqueta/colorante fluorescente adecuada), un cromóforo (por ejemplo, biotina, digoxigenina y similares), puntos cuánticos, o partículas de oro. El marcador detectable puede unirse por medios convencionales a cualquier aminoácido de la proteína.

ARN de guía. El sistema de nucleasa CRISPR/Cas también comprende un ARN de guía (ARNg). El ARN de guía interactúa con la nucleasa CRISPR/Cas y el sitio diana para guiar la nucleasa CRISPR/Cas al sitio diana en la secuencia cromosómica. El sitio diana no tiene limitación de secuencia, excepto que la secuencia está bordeada por un motivo adyacente protospacer (PAM, forma siglada de *protospacer adjacent motif*). Por ejemplo, las secuencias PAM para las proteínas Cas9 incluyen 3'-NGG, 3'-NGGNG, 3'-NNAGAAW y 3'-ACAY, y las secuencias PAM para Cpf1 incluyen 5'-TTN (en las que N se define como cualquier nucleótido, W se define como A o T, e Y se define como C o T).

Cada ARN de guía puede comprender tres regiones: una primera región en el extremo 5' que tiene complementariedad con el sitio diana en la secuencia de ADN cromosómico, una segunda región que es interna y forma una estructura de tallo-bucle, y una tercera región en el extremo 3' que permanece esencialmente monocatenaria. Las segunda y tercera regiones forman una estructura secundaria que interactúa con la proteína CRISPR/Cas. La primera región de cada ARN de guía es distinta (es decir, es de secuencia específica). Las segunda y tercera regiones pueden ser las mismas en los ARN de guía que forman complejo con una proteína CRISPR/Cas particular.

La primera región del ARN de guía tiene complementariedad con la secuencia (es decir, la secuencia del protospacer) en el sitio diana, de forma que la primera región del ARN de guía pueda emparejarse con la secuencia diana. Por ejemplo, la primera región de un ARN de guía de SpCas9 puede comprender GN<sub>17-20</sub>GG. En general, la complementariedad entre la primera región (es decir, el ARNcr) del ARN de guía y la secuencia diana es al menos del 80 %, al menos del 85 %, al menos del 90 %, al menos del 95 % o más. En diversas realizaciones, la primera región del ARN de guía puede comprender de aproximadamente 10 nucleótidos a más de aproximadamente 25 nucleótidos. Por ejemplo, la región de emparejamiento de bases entre la primera región del ARN de guía y el sitio diana en la secuencia de ADNc puede ser de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25 o de más de 25 nucleótidos de longitud. En una realización ejemplar, la primera región del ARN de guía es de aproximadamente 19, 20 o 21 nucleótidos de longitud.

El ARN de guía también comprende una segunda región que forma una estructura secundaria. En algunas realizaciones, la estructura secundaria comprende al menos un tallo (u horquilla) y un bucle. La longitud de cada bucle y el tallo puede variar. Por ejemplo, el bucle puede variar de aproximadamente 3 a aproximadamente 10 nucleótidos de longitud, y el tallo puede variar de aproximadamente 6 a aproximadamente 20 pares de bases de longitud. El tallo puede comprender uno o más protuberancias de 1 a aproximadamente 10 nucleótidos. Por lo tanto, la longitud total de la segunda región puede variar de aproximadamente 16 a aproximadamente 60 nucleótidos de longitud. El ARN de guía también comprende una tercera región en el extremo 3' que permanece esencialmente monocatenaria. Por lo tanto, la tercera región no tiene complementariedad con ninguna secuencia de ácido nucleico en la célula de interés y no tiene complementariedad con el resto del ARN de guía. La longitud de la tercera región puede variar. En general, la tercera región es de más de aproximadamente 4 nucleótidos de longitud. Por ejemplo, la longitud de la tercera región puede variar de aproximadamente 5 a aproximadamente 60 nucleótidos de longitud.

La longitud combinada de las segunda y tercera regiones (también llamada región universal o de armazón) del ARN de guía puede variar de aproximadamente 30 a aproximadamente 120 nucleótidos de longitud. En un aspecto, la longitud combinada de las segunda y tercera regiones del ARN de guía puede variar de aproximadamente 70 a aproximadamente 100 nucleótidos de longitud.

5 En aún otras realizaciones, las segunda y tercera regiones del ARN de guía pueden comprender una o más regiones adicionales de tallo-bucle, en las que las regiones de tallo-bucle comprenden secuencias aptaméricas (Konerermann y col., Nature3, 2015, 517(7536):583-588; Zalatan y col., Cell, 2015, 160(1-2):339-50). Las secuencias aptaméricas adecuadas incluyen las que se unen a proteínas adaptadoras elegidas de MS2, PP7, COM, Q $\beta$ , F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205,  $\phi$ Cb5,  $\phi$ Cb8r,  $\phi$ Cb12r,  $\phi$ Cb23r, 7s, PRR1, HSF1, AID, APOBEC1, p300, TET1/2/3, VP64, GFP, Rta, p65, MyoD1 o VP160. En tales realizaciones, la longitud total de la segunda y tercera regiones del ARN de guía puede variar de hasta aproximadamente 125 nucleótidos, hasta aproximadamente 150 nucleótidos, hasta aproximadamente 175 nucleótidos, hasta aproximadamente 200 nucleótidos, hasta aproximadamente 225 nucleótidos, hasta aproximadamente 250 nucleótidos, hasta aproximadamente 275 nucleótidos, o hasta aproximadamente 300 nucleótidos.

En algunas realizaciones, el ARN de guía puede ser una molécula única que comprende las tres regiones. En otras realizaciones, el ARN de guía puede comprender dos moléculas separadas. La primera molécula de ARN (es decir, ARNcr) puede comprender la primera región del ARN de guía y la mitad del "tallo" de la segunda región del ARN de guía. La segunda molécula de ARN (es decir, ARNtracr) puede comprender la otra mitad del "tallo" de la segunda región del ARN de guía y la tercera región del ARN de guía. Por lo tanto, en esta realización, las primera y segunda moléculas de ARN contienen cada una una secuencia de nucleótidos que son complementarias entre sí. Por ejemplo, en una realización, las moléculas de ARNcr y ARNtracr comprenden cada una una secuencia (de aproximadamente 6 a aproximadamente 20 nucleótidos) que se empareja con la otra secuencia para formar un ARN de guía funcional. Por ejemplo, el ARN de guía de los sistemas CRISPR/Cas de tipo II puede comprender ARNcr y ARNtracr. En algunos aspectos, el ARNcr para un sistema CRISPR/Cas de tipo II puede sintetizarse químicamente y el sistema CRISPR/Cas de tipo II de ARNtracr puede sintetizarse *in vitro* (véase la sección (I)(c) a continuación). En otras realizaciones, el ARN de guía de los sistemas CRISPR/Cas de tipo V puede comprender solo ARNcr.

El ARN de guía puede comprender ribonucleótidos convencionales, ribonucleótidos modificados (por ejemplo, pseudouridina), isómeros de ribonucleótidos y/o análogos de ribonucleótidos. En algunas realizaciones, el ARN de guía puede comprender adicionalmente al menos un marcador detectable. El marcador detectable puede ser un fluoróforo (por ejemplo, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluor, etiquetas Halo o un colorante fluorescente adecuado), un cromóforo (por ejemplo, biotina, digoxigenina y similares), puntos cuánticos, o partículas de oro. Los expertos en la materia están familiarizados con el diseño y construcción del ARNg, por ejemplo, las herramientas de diseño de ARNg están disponibles en la Internet o de fuentes comerciales.

El ARN de guía puede sintetizarse químicamente, sintetizarse enzimáticamente, o una combinación de los mismos. Por ejemplo, el ARN de guía se puede sintetizar utilizando procedimientos convencionales de síntesis en fase sólida basados en fosoramidita. Como alternativa, el ARN de guía se puede sintetizar *in vitro* uniendo operativamente el ADN que codifica el ARN de guía a una secuencia de control de promotor reconocida por una ARN polimerasa de fago. Los ejemplos de secuencias de promotor de fagos adecuadas incluyen las secuencias de promotor de T7, T3, SP6, o variaciones de las mismas. En realizaciones en las que el ARN de guía comprende dos moléculas distintas (es decir, ARNcr y ARNtracr), el ARNcr puede sintetizarse químicamente y el ARNtracr puede sintetizarse enzimáticamente. El ácido nucleico que codifica el ARN de guía puede ser parte de un vector plasmídico, que puede comprender adicionalmente secuencias de control de la expresión adicionales (por ejemplo, secuencias potenciadoras, secuencias de Kozak, secuencias de poliadenilación, secuencias de terminación de la transcripción, etc.), secuencias de marcadores de selección (por ejemplo, genes de resistencia a antibióticos), orígenes de replicación, y similares. Como se detalla a continuación en la sección (I)(c), el ácido nucleico que codifica el ARN de guía puede unirse operativamente a una secuencia de control de promotor que es reconocida por la ARN polimerasa III (Pol III), para su expresión en células eucariotas.

#### (ii) Sistemas de nickasa doble CRISPR/Cas

En otras realizaciones, la proteína de modificación de ADN programable puede ser un sistema de nickasa doble CRISPR/Cas. Los sistemas de nickasa doble CRISPR/Cas son similares a los sistemas de nucleasa CRISPR/Cas descritos anteriormente en la sección (I)(a)(i), excepto que la nucleasa CRISPR/Cas se modifica para que escinda solo una cadena de ADN. Por lo tanto, un sistema de nickasa simple CRISPR/Cas crea una rotura monocatenaria o una muesca en el ADN bicatenario, y un sistema de nickasa doble CRISPR/Cas emparejado que comprende ARN de guía desplazados emparejados crea una ruptura bicatenaria en el ADN.

Una nucleasa CRISPR/Cas puede convertirse en una nickasa mediante una o más mutaciones y/o deleciones. Por ejemplo, una Cas9 nickase puede comprender una o más mutaciones en uno de los dominios de nucleasa (por ejemplo, el dominio de tipo RuvC o el dominio de tipo HNH). Por ejemplo, la una o más mutaciones pueden ser D10A, D8A, E762A y/o D986A en el dominio tipo RuvC o la una o más mutaciones pueden ser H840A, H559A, N854A, N856A y/o N863A en el dominio tipo HNH.

**(b) Proteínas de unión a ADN programables**

La composición también comprende al menos una proteína de unión a ADN programable. Las proteínas de unión a ADN programables son proteínas que se unen a secuencias de ADN específicas pero no modifican el ADN o la proteína (o proteínas) asociada con el ADN.

- 5 La al menos una proteína de unión a ADN programable es una nucleasa CRISPR/Cas modificada para carecer de actividad nucleasa. Por ejemplo, la proteína de unión a ADN programable puede ser un sistema CRISPR/Cas catalíticamente inactivo. Para ello, la nucleasa CRISPR/Cas puede modificarse mediante mutación y/o delección para eliminar toda la actividad nucleasa. En una realización, tanto el dominio de tipo RuvC como el dominio de tipo HNH comprenden una o más mutaciones y/o delecciones para eliminar la actividad nucleasa. Por ejemplo, la proteína CRISPR/Cas catalíticamente inactiva puede ser una Cas9 catalíticamente inactiva (muerta) (dCas9) en la que el dominio de tipo RuvC comprende una mutación D10A, D8A, E762A y/o D986A y el dominio de tipo HNH comprende una mutación H840A, H559A, N854A, N856A y/o N863A. Como alternativa, la proteína CRISPR/Cas catalíticamente inactiva puede ser una proteína Cpf1 catalíticamente inactiva (muerta) que comprende mutaciones comparables en el dominio de nucleasa. En otros aspectos, la proteína de unión a ADN programable puede ser una proteína CRISPR/Cas modificada para mellar una cadena de una secuencia bicatenaria (es decir, es una como se detalla a anteriormente en la sección (I)(a)(ii).

La proteína de unión a ADN programable puede comprender también al menos una señal de localización nuclear, un dominio de penetración celular, un dominio marcador y/o una etiqueta detectable, los cuales se describen anteriormente en la sección (I)(a)(i).

**20 (c) Ácidos nucleicos que codifican proteínas de modificación de ADN programables o proteínas de unión a ADN programables**

- El ácido nucleico que codifica la proteína de modificación de ADN programable, descrito anteriormente en la sección (I)(a), o la proteína de unión a ADN programable, descrito anteriormente en la sección (I)(b), puede ser ADN o ARN, lineal o circular, monocatenario o bicatenario. El ARN o el ADN puede tener codones optimizados para una traducción eficaz a proteínas en la célula eucariota de interés. Los programas de optimización de codones están disponibles como software gratuito o de fuentes comerciales.

- En algunas realizaciones, el ácido nucleico que codifica la proteína de modificación de ADN programable o la al menos una proteína de unión a ADN programable puede ser ARNm. El ARNm se puede sintetizar *in vitro*. Para ello, el ADN que codifica la proteína de modificación de ADN o la al menos una proteína de unión a ADN puede unirse operativamente a una secuencia de promotor que es reconocida por una ARN polimerasa de fago para la síntesis *in vitro* de ARNm. Por ejemplo, la secuencia de promotor puede ser una secuencia de promotor de T7, T3 o SP6 o una variación de una secuencia de promotor de T7, T3 o SP6. En tales realizaciones, el ARN transcrito *in vitro* puede purificarse, protegerse de forma terminal y/o poliadenilarse. Como se detalla a continuación, el ADN que codifica la proteína de modificación de ADN o la proteína de unión a ADN puede ser parte de un vector.

- 35 En otras realizaciones, el ácido nucleico que codifica la proteína de modificación de ADN programable o la al menos una proteína de unión a ADN programable puede ser ADN. La secuencia de ADN que codifica la proteína de modificación de ADN programable o la al menos una proteína de unión a ADN programable puede unirse operativamente a al menos una secuencia de control de promotor para la expresión en la célula de interés. En algunas realizaciones, la secuencia de ADN codificante también se puede unir a una señal de poliadenilación (por ejemplo, señal poliA de SV40, señal poliA de la hormona de crecimiento bovina (BGH), *etc.*) y/o al menos una secuencia de terminación de la transcripción.

- En determinadas realizaciones, la secuencia de ADN codificante puede estar operativamente unida a una secuencia de promotor para la expresión de la proteína de modificación de ADN o la proteína de unión a ADN en células bacterianas (por ejemplo, *E. coli*) o células eucariotas (por ejemplo, levaduras, de insecto o de mamífero). Los promotores bacterianos adecuados incluyen, sin limitación, los promotores de T7, los promotores de operón *lac*, los promotores *trp*, los promotores *tac* (que son híbridos de promotores *trp* y *lac*), variaciones de cualquiera de los anteriores, y combinaciones de cualquiera de los anteriores. Los ejemplos no limitantes de promotores eucariotas adecuados incluyen promotores constitutivos, regulados o específicos de células o tejidos. Las secuencias de control de promotores constitutivos eucariotas adecuadas incluyen, pero sin limitación, el promotor temprano inmediato de citomegalovirus (CMV), el promotor del virus de simio (SV40), el promotor tardío principal de adenovirus, el promotor del virus del sarcoma de Rous (VSR), el promotor del virus del tumor mamario de ratón (VTMR), el promotor de la fosfoglicerato quinasa (PGK), el promotor del factor de elongación (ED1) alfa, promotores de ubiquitina, promotores de actina, promotores de tubulina, promotores de inmunoglobulina, fragmentos de los mismos, o combinaciones de cualquiera de los anteriores. Los ejemplos de secuencias de control de promotor reguladas eucariotas adecuadas incluyen, sin limitación, las reguladas por choque térmico, metales, esteroides, antibióticos o alcohol. Los ejemplos no limitantes de promotores específicos de tejido incluyen el promotor de B29, el promotor de CD14, el promotor de CD43, el promotor de CD45, el promotor de CD68, el promotor de desmina, el promotor de elastasa 1, el promotor de endoglin, el promotor de fibronectina, el promotor de Flt-1, el promotor de GFAP, el promotor de GPIIb, el promotor de ICAM-2, el promotor de INF- $\beta$ , el promotor de Mb, el promotor de Nphs1, el promotor de OG-2, el promotor de SP-

B, el promotor de SYN1 y el promotor de WASP. La secuencia de promotora puede ser de tipo silvestre o puede modificarse para una expresión más eficiente o eficaz.

En diversas realizaciones, el ácido nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable puede estar presente en un vector. Los vectores adecuados incluyen vectores plasmídicos, fagémidos, cósmidos, cromosomas artificiales/mini-cromosomas, transposones y vectores víricos (por ejemplo, vectores lentivíricos, vectores víricos adenoasociados, vectores adenovíricos, etc.). En una realización, el ADN nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable puede estar presente en un vector plasmídico. Los ejemplos no limitantes de vectores plasmídicos adecuados incluyen pUC, pBR322, pET, pBluescript y variantes de los mismos. En otras realizaciones, el ácido nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable puede estar presente en un vector vírico. El plásmido o vector vírico puede comprender secuencias de control de la expresión adicionales (por ejemplo, secuencias potenciadoras, secuencias de Kozak, secuencias de poliadenilación, secuencias de terminación de la transcripción, etc.), secuencias de marcadores de selección (por ejemplo, genes de resistencia a antibióticos), orígenes de replicación, y similares. Se puede encontrar información adicional en "Current Protocols in Molecular Biology" Ausubel y col., John Wiley & Sons, Nueva York, 2003 o "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3ª edición, 2001.

En realizaciones en que la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable comprende una proteína CRISPR/Cas o una variante de la misma, el vector de expresión que comprende el ácido nucleico que codifica la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable puede comprender adicionalmente una secuencia que codifica uno o más ARN de guía. La secuencia que codifica el ARN de guía generalmente está operativamente unida a al menos una secuencia de control de la transcripción para la expresión del ARN de guía en la célula eucariota de interés. Por ejemplo, el ácido nucleico que codifica el ARN de guía puede unirse operativamente a una secuencia de promotor que sea reconocida por la ARN polimerasa III (Pol III). Los ejemplos de promotores Pol III incluyen, pero sin limitación, los promotores de mamífero de ARN U6, U3, H1 y 7SL.

**(d) Composiciones específicas**

En algunas realizaciones, la proteína de modificación de ADN programable y las una o más proteínas de unión a ADN programables se proporcionan como proteínas (o, en algunos casos, como complejos de proteína-ARN). Las proteínas de modificación de ADN programables y las proteínas de unión a ADN programables pueden expresarse en células bacterianas o eucariotas, y purificarse usando medios bien conocidos en la técnica. En otras realizaciones, la proteína de modificación de ADN programable y las una o más proteínas de unión a ADN programables se proporcionan como ácidos nucleicos.

En algunas realizaciones, la composición puede comprender una proteína/sistema de unión a ADN programable o ácidos nucleicos codificantes. En otras realizaciones, la composición puede comprender dos proteínas/sistemas de unión a ADN programables o ácidos nucleicos codificantes. Aún en otras realizaciones, la composición puede comprender tres proteínas/sistemas de unión a ADN programables o ácidos nucleicos codificantes. En realizaciones adicionales, la composición puede comprender cuatro proteínas/sistemas de unión a ADN programables o ácidos nucleicos codificantes. En aún otras realizaciones, la composición puede comprender cinco o más o proteínas/sistemas de unión a ADN programables o ácidos nucleicos codificantes.

En realizaciones específicas, la proteína de modificación de ADN programable puede comprender un sistema CRISPR/Cas (por ejemplo, nucleasa CRISPR/Cas, nickasa doble CRISPR/Cas, o proteína CRISPR/Cas catalíticamente inactiva (muerta) unida a un dominio de modificación no nucleasa) y la proteína de unión a ADN programable puede ser un sistema CRISPR/Cas que carece de actividad nucleasa. Por ejemplo, la proteína de unión a ADN programable puede ser un sistema CRISPR/Cas catalíticamente inactivo. En general, cada proteína CRISPR/Cas comprende al menos una señal de localización nuclear. En algunas iteraciones, la composición puede comprender los sistemas CRISPR/Cas como proteínas CRISPR/Cas y ARN de guía, en la que la proteína y el ARN pueden ser entidades distintas o la proteína y el ARN pueden formar complejos entre sí. El ARN de guía puede sintetizarse químicamente al menos de forma parcial. El ARN de guía puede sintetizarse enzimáticamente. En otras iteraciones, la composición puede comprender las proteínas CRISPR/Cas y el ADN que codifican los ARN de guía. En aún otras iteraciones, la composición puede comprender ARNm que codifica las proteínas CRISPR/Cas y el ADN que codifican los ARN de guía. En todavía otras iteraciones, la composición puede comprender vectores plasmídicos o víricos que codifican las proteínas CRISPR/Cas y/o los ARN de guía. En determinadas realizaciones, la proteína CRISPR/Cas catalíticamente activa y la proteína CRISPR/Cas catalíticamente inactiva (muerta) son proteínas Cas9. Los ácidos nucleicos que codifican las proteínas CRISPR/Cas tienen, generalmente, codones optimizados para una expresión óptima en la célula eucariota de interés.

**(II) Kits**

Un aspecto adicional de la presente divulgación proporciona kits que comprenden las composiciones detalladas anteriormente en la sección (I). Los kits pueden proporcionar la proteína de modificación de ADN programable o la al

menos una de las proteínas de unión a ADN programable como proteínas, como complejos de proteína-ARN o como ácidos nucleicos que codifican los diversos componentes, como se detalla anteriormente. Los kits pueden comprender adicionalmente reactivos de transfección, medios de cultivo celular, medios de selección, reactivos de transcripción *in vitro*, reactivos de purificación de ácido nucleico, reactivos de purificación de proteínas, tampones, y similares. Los kits proporcionados en el presente documento generalmente incluyen instrucciones para llevar a cabo los procedimientos detallados a continuación. Las instrucciones incluidas en los kits pueden fijarse al material de embalaje o pueden incluirse como un prospecto de envase. Si bien las instrucciones suelen ser materiales escritos o impresos, no se limitan a tales. La presente divulgación contempla cualquier medio capaz de almacenar tales instrucciones y comunicarlas a un usuario final. Dichos medios incluyen, pero sin limitación, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, CD ROM) y similares. Como se usa en el presente documento, el término "instrucciones" puede incluir la dirección de un sitio de Internet que proporciona las instrucciones.

En algunas realizaciones, la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable del kit puede comprender un sistema CRISPR/Cas de tipo II. En determinadas realizaciones, el ARN de guía del sistema CRISPR/Cas de tipo II puede comprender ARNcr y ARNtracr. El kit, por lo tanto, puede proporcionar el ARNtracr (o los ARNtracr) universal, y el usuario final del kit puede proporcionar el ARNcr (o los ARNcr) específico de secuencia. En algunos aspectos, el kit puede comprender la proteína (o proteínas) CRISPR/Cas de tipo II y el ARNtracr (o los ARNtracr). En otros aspectos, el kit puede comprender ARNm o ADN que codifica la proteína (o proteínas) CRISPR/Cas de tipo II y ADN que codifica el ARNtracr (o los ARNtracr).

En aún otras realizaciones, la proteína de modificación de ADN programable y/o la al menos una proteína de unión a ADN programable del kit puede comprender un sistema CRISPR/Cas de tipo V. Como se detalla anteriormente, el ARN de guía de los sistemas CRISPR/Cas de tipo V comprende solo ARNcr. En algunos aspectos, el kit puede comprender la proteína (o proteínas) de CRISPR/Cas de tipo V y el ARNcr (o los ARNcr), o el kit puede comprender ARNm o ADN que codifica la proteína (o proteínas) de CRISPR/Cas de tipo V y ADN que codifica el ARNcr (o los ARNcr). En otros aspectos, el kit puede comprender solo la proteína (o proteínas) CRISPR/Cas de tipo V o un ácido nucleico que codifica la proteína (o proteínas) CRISPR/Cas de tipo V, en el que el usuario final del kit proporciona el ARNcr (o los ARNcr).

**(III) Procedimientos para aumentar la accesibilidad a sitios cromosómicos que se tienen como objetivo**

Otro aspecto de la presente divulgación abarca procedimientos para aumentar la eficacia y/o la especificidad de la modificación epigenética/genoma dirigida en células eucariotas, al aumentar la accesibilidad de una proteína de modificación de ADN programable a su secuencia diana en el ADN cromosómico. Los procedimientos comprenden introducir en la célula eucariota de interés (a) una proteína de modificación de ADN programable o un ácido nucleico que codifica la proteína de modificación de ADN programable y (a) al menos una proteína de unión a ADN programable o un ácido nucleico que codifica la al menos una proteína de unión a ADN programable. La proteína de modificación de ADN programable se genomanipula para que reconozca y se una a una secuencia diana en el ADN cromosómico, sitio en el que la proteína de modificación de ADN puede modificar el ADN o la proteína (o proteínas) asociada. Cada una de las una o más proteínas de unión a ADN programables se genomanipula para reconocer y unirse a una secuencia próxima a la secuencia cromosómica diana de la proteína de modificación de ADN. Las proteínas de modificación de ADN programables y las proteínas de unión a ADN programables se detallan anteriormente en la sección (I).

En general, la secuencia próxima a la secuencia cromosómica diana está ubicada dentro de aproximadamente 250 pares de bases a cada lado (es decir, corriente arriba o corriente abajo) de la secuencia cromosómica diana. El sitio (o sitios) próximo puede ubicarse en cualquiera de las cadenas del ADN dúplex. En algunas realizaciones, la secuencia próxima a la secuencia cromosómica diana puede ubicarse a menos de aproximadamente 250 pb, menos de aproximadamente 200 pb, menos de aproximadamente 150 pb, menos de aproximadamente 100 pb, menos de aproximadamente 75 pb, menos de aproximadamente 50 pb, menos de aproximadamente 25 pb, menos de aproximadamente 20 pb, menos de aproximadamente 15 pb, menos de aproximadamente 10 pb o menos de aproximadamente 5 pb de la secuencia cromosómica diana de la proteína de modificación de ADN. En determinadas realizaciones, la secuencia próxima a la secuencia cromosómica diana puede ubicarse a menos de aproximadamente 1 pb a aproximadamente 10 pb, de aproximadamente 11 pb a aproximadamente 20 pb, de aproximadamente 21 pb a aproximadamente 30 pb, de aproximadamente 31 pb a aproximadamente 40 pb, de aproximadamente 41 pb a aproximadamente 50 pb, de aproximadamente 51 pb a aproximadamente 60 pb, de aproximadamente 61 pb a aproximadamente 70 pb, de aproximadamente 71 pb a aproximadamente 80 pb, de aproximadamente 81 pb a aproximadamente 90 pb, de aproximadamente 91 pb a aproximadamente 100 pb, de aproximadamente 101 pb a aproximadamente 150 pb, de aproximadamente 151 pb a aproximadamente 200 pb, o de aproximadamente 201 pb a aproximadamente 250 pb a cada lado de la secuencia cromosómica diana. En otras realizaciones, la secuencia próxima a la secuencia cromosómica diana puede ubicarse a menos de aproximadamente 5 pb a aproximadamente 75 pb, de aproximadamente 10 pb a aproximadamente 50 pb, o de aproximadamente 15 pb a aproximadamente 25 pb a cada lado de la secuencia cromosómica diana.

En algunas realizaciones, el procedimiento comprende introducir en la célula al menos una proteína de unión a ADN programable cuya secuencia de unión está ubicada corriente arriba o corriente abajo de la secuencia cromosómica

diana. En otras realizaciones, el procedimiento comprende introducir en la célula al menos dos proteínas de unión a ADN programables, en las que la secuencia de unión de una está ubicada corriente arriba de la secuencia cromosómica diana y la secuencia de unión de la otra está ubicada corriente abajo de la secuencia cromosómica diana. En realizaciones adicionales, el procedimiento comprende introducir en la célula al menos tres proteínas de unión a ADN programables cuyas secuencias de unión están ubicadas corriente arriba o corriente abajo de la secuencia cromosómica diana. En realizaciones adicionales, el procedimiento comprende introducir en la célula al menos cuatro o más proteínas de unión a ADN programables cuyas secuencias de unión están ubicadas corriente arriba o corriente abajo de la secuencia cromosómica diana. En estas realizaciones, por ejemplo, el procedimiento puede comprender introducir una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más de diez proteínas de unión a ADN programables cuyas secuencias de unión se ubican dentro de aproximadamente 250 pb a cada lado (es decir, corriente arriba o corriente abajo) de la secuencia cromosómica diana.

La unión de cada una de las una o más proteínas de unión a ADN programables al sitio próximo a la secuencia cromosómica diana cambia la configuración local de la cromatina, lo que conduce a un aumento de la accesibilidad de la proteína de modificación de ADN programable a la secuencia cromosómica diana (previamente inaccesible) (véase la FIG. 1). Como consecuencia, la eficacia de la modificación por la proteína de modificación de ADN aumenta (véanse, por ejemplo, los Ejemplos 1-3). Dicho de otra manera, la eficacia de la modificación por una proteína de modificación de ADN aumenta cuando la proteína de modificación de ADN se introduce en la célula en combinación con una o más proteínas de unión a ADN programables, en comparación con cuando la proteína de modificación de ADN se introduce en la célula sola.

Además, los procedimientos desvelados en el presente documento aumentan la especificidad de la modificación del genoma dirigida. Aunque la proteína de modificación de ADN programable se genomanipula para que reconozca y se una a una secuencia diana en un locus cromosómico específico, pueden existir secuencias idénticas o casi idénticas en otras ubicaciones cromosómicas (lo que da resultado efectos inespecíficos). En realizaciones en que la unión de una proteína de modificación de ADN programable a una secuencia cromosómica diana depende en gran medida de la unión de una o más proteínas de unión a ADN programables a secuencias próximas a la secuencia cromosómica diana, la unión de las una o más proteínas de unión a ADN programables al sitio (o sitios) próximo a la secuencia diana en el locus cromosómico de interés, sin embargo, proporciona especificidad adicional al suceso de modificación (véase el Ejemplo 4).

Por lo tanto, los procedimientos desvelados en el presente documento pueden aumentar la especificidad de la edición dirigida del genoma (por ejemplo, correcciones génicas, genosupresión génica, genoinserción, y similares), modificaciones epigenéticas dirigidas y regulación transcripcional dirigida.

#### **(a) Introducción en la célula**

Como se ha descrito, el procedimiento comprende introducir en la célula (a) una proteína de modificación de ADN programable o un ácido nucleico que codifica la proteína de modificación de ADN programable y (b) al menos una proteína de unión a ADN programable o un ácido nucleico que codifica la al menos una proteína de unión a ADN programable. Las proteínas de modificación de ADN programables se detallan anteriormente en la sección (I)(a), las proteínas de unión a ADN programables se detallan anteriormente en la sección (I)(b), y los ácidos nucleicos que codifican las proteínas de modificación de ADN o la proteína de unión a ADN programable se describen anteriormente en la sección (I)(c).

La proteína de modificación de ADN programable o el ácido nucleico que codifica la proteína de modificación de ADN programable y la al menos una proteína de unión a ADN programable o el ácido nucleico que codifica la al menos una proteína de unión a ADN programable se pueden introducir en la célula de interés por una diversidad de medios.

En algunas realizaciones, la célula puede transfectarse con las moléculas apropiadas (es decir, proteína, ADN y/o ARN). Los procedimientos de transfección adecuados incluyen nucleofección (o electroporación), transfección mediada por fosfato de calcio, transfección de polímeros catiónicos (por ejemplo, DEAE-dextrano o polietilénimina), transducción vírica, transfección de virosomas, transfección de viriones, transfección de liposomas, transfección de liposomas catiónicos, transfección de inmunoliposomas, transfección de lípidos no liposómicos, transfección de dendrímeros, transfección por choque térmico, magnetofección, lipofección, suministro por pistola de genes, impalefacción, sonoporación, transfección óptica y captación de ácidos nucleicos potenciada por agentes patentados. Los procedimientos de transfección son bien conocidos en la técnica. (véase, por ejemplo, "Current Protocols in Molecular Biology" Ausubel y col., John Wiley & Sons, Nueva York, 2003 o "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3ª edición, 2001). En otras realizaciones, las moléculas pueden introducirse en la célula mediante microinyección. Por ejemplo, las moléculas pueden inyectarse en el citoplasma o los núcleos de las células de interés. La cantidad de cada molécula introducida en la célula puede variar, pero los expertos en la materia están familiarizados con los medios para determinar la cantidad apropiada.

Las diversas moléculas pueden introducirse en la célula de forma simultánea o secuencial. Por ejemplo, la proteína de modificación de ADN programable (o su ácido nucleico codificante) y la al menos una proteína de unión a ADN programable (o ácido nucleico codificante) pueden introducirse al mismo tiempo. Como alternativa, se puede introducir en la célula primero uno y más tarde se puede introducir el otro.

En general, la célula se mantiene en condiciones apropiadas para el crecimiento y/o mantenimiento celular. Las condiciones adecuadas de cultivo celular son bien conocidas en la técnica y se describen, por ejemplo, en Santiago y col., Proc. Natl. Acad. Sci. USA, 2008, 105:5809-5814; Moehle y col. Proc. Natl. Acad. Sci. USA, 2007, 104:3055-3060; Urnov y col., Nature, 2005, 435:646-651; y Lombardo y col., Nat. Biotechnol., 2007, 25:1298-1306. Los expertos en la materia aprecian que los procedimientos para el cultivo de células son conocidos en la técnica y pueden variar y variarán según el tipo de célula. Se puede usar una optimización de rutina, en todos los casos, para determinar las mejores técnicas para un tipo de célula particular.

**(b) Modificación dirigida del genoma**

La unión de las una o más proteínas de unión a ADN programables a la secuencia (o secuencias) próxima a la secuencia cromosómica diana cambia la configuración local de la cromatina, por ejemplo, la estructura nucleosómica puede alterarse y/o las histonas pueden desplazarse. Como consecuencia, la proteína de modificación de ADN programable puede acceder mejor a la secuencia cromosómica diana, en comparación con cuando la proteína de modificación de ADN programable se usa sola. La accesibilidad aumentada da como resultado una eficacia y/o especificidad aumentadas de la modificación del genoma dirigida. La modificación epigenética/genómica dirigida puede estar mediada por proteínas de modificación de ADN que tienen actividad de nucleasa o no tienen actividad de nucleasa.

En realizaciones en las que la proteína de modificación de ADN programable tiene actividad de nucleasa, la proteína de modificación del ADN puede introducir una rotura bicatenaria en la secuencia cromosómica diana. La rotura bicatenaria en la secuencia cromosómica puede repararse mediante un proceso de reparación de unión de extremos no homólogos (NHEJ). Debido a que el NHEJ es propenso a errores, pueden producirse deleciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido, o combinaciones de los mismos, durante la reparación de la rotura. Por consiguiente, la secuencia cromosómica que se tiene como objetivo puede modificarse o desactivarse. Por ejemplo, una eliminación, inserción o sustitución en el cambio en el marco de lectura de una secuencia codificante puede conducir a un producto proteico alterado, o a ningún producto proteico (lo que se denomina "genosuprimido"). En algunas iteraciones, el procedimiento puede comprender adicionalmente introducir en la célula un polinucleótido donante (véase más abajo) que comprende una secuencia donante que está flanqueada por una secuencia que tiene una identidad de secuencia sustancial con las secuencias ubicadas a cada lado de la secuencia cromosómica diana, de forma que durante la reparación de la rotura bicatenaria mediante un proceso de reparación dirigido por homología (HDR), la secuencia del donante en el polinucleótido del donante puede intercambiarse o integrarse en la secuencia cromosómica en la secuencia cromosómica diana. La integración de una secuencia exógena se denomina "genoinserción". Como se detalla anteriormente, los procedimientos desvelados en el presente documento también reducen los efectos inespecíficos, aumentando de este modo la especificidad de la modificación del genoma dirigida.

En diversas iteraciones, por lo tanto, la eficacia y/o la especificidad de la modificación dirigida del genoma puede aumentarse en al menos aproximadamente 0,1 veces, al menos aproximadamente 0,5 veces, al menos aproximadamente 1 veces, al menos aproximadamente 2 veces, al menos aproximadamente 5 veces, al menos aproximadamente 10 veces o al menos aproximadamente 20 veces, al menos aproximadamente 50 veces, al menos aproximadamente 100 veces, o más de aproximadamente 100 veces con respecto a cuando la proteína de modificación de ADN programable que tiene actividad de nucleasa se usa sola. Por ejemplo, la proteína de modificación de ADN programable que tiene actividad de nucleasa, cuando se usa sola, no puede tener detectables o sucesos de integración indels. Sin embargo, cuando la proteína de modificación de ADN programable que tiene actividad de nucleasa se usa en combinación con al menos una proteína de unión a ADN programable, se pueden detectar indels y sucesos de integración (por ejemplo, al menos aproximadamente el 1 % de indels/integraciones, al menos aproximadamente el 5 % de indels/integración, al menos aproximadamente el 10 % de indels/integraciones, al menos aproximadamente el 20 % de indels/integraciones, al menos aproximadamente el 30 % de indels/integraciones, al menos aproximadamente el 40 % de indels/integraciones, al menos aproximadamente el 50 % de indels/integraciones, o más de aproximadamente el 50 % de indels/integraciones).

Las modificaciones del genoma dirigidas detalladas anteriormente se pueden realizar de forma individual o multiplexada (es decir, pueden tenerse como objetivo dos o más secuencias cromosómicas de forma simultánea).

**(c) Polinucleótido donante opcional**

En realizaciones en las que la proteína de modificación de ADN programable comprende actividad de nucleasa, el procedimiento puede comprender adicionalmente introducir en la célula al menos un polinucleótido donante. El polinucleótido donante puede ser monocatenario o bicatenario, lineal o circular, y/o ARN o ADN. En algunas realizaciones, el polinucleótido donante puede ser un vector, por ejemplo, un vector plasmídico.

El polinucleótido donante comprende al menos una secuencia donante. En algunos aspectos, la secuencia donante del polinucleótido donante puede ser una versión modificada de una secuencia cromosómica endógena o nativa. Por ejemplo, la secuencia donante puede ser esencialmente idéntica a una porción de la secuencia cromosómica en o cerca de la secuencia que tiene como diana la proteína de modificación de ADN, pero comprender al menos un cambio de nucleótido. Por lo tanto, tras la integración o el intercambio con la secuencia nativa, la secuencia en la ubicación

5 cromosómica que es el objetivo comprende al menos un cambio de nucleótido. Por ejemplo, el cambio puede ser una inserción de uno o más nucleótidos, una delección de uno o más nucleótidos, una sustitución de uno o más nucleótidos, o combinaciones de los mismos. Como consecuencia de la integración de "corrección génica" de la secuencia modificada, la célula puede producir un producto genético modificado a partir de la secuencia cromosómica que se tiene como objetivo.

10 En otros aspectos, la secuencia donante del polinucleótido donante puede ser una secuencia exógena. Como se usa en el presente documento, una secuencia "exógena" se refiere a una secuencia que no es nativa de la célula, o una secuencia cuya ubicación nativa está en una ubicación distinta en el genoma de la célula. Por ejemplo, la secuencia exógena puede comprender una secuencia codificante de proteínas, que se puede unir operativamente a una secuencia de control de promotor exógena de forma que, tras la integración en el genoma, la célula puede expresar la proteína codificada por la secuencia integrada. Como alternativa, la secuencia exógena puede integrarse en la secuencia cromosómica de forma que su expresión esté regulada por una secuencia de control de promotor endógena. En otras iteraciones, la secuencia exógena puede ser una secuencia de control de la transcripción, otra secuencia de control de la expresión, una secuencia de codificación de ARN, y así sucesivamente. Como se ha indicado anteriormente, la integración de una secuencia exógena en una secuencia cromosómica se denomina "genoinserción".

15 Como pueden apreciar los expertos en la materia, la longitud de la secuencia del donante puede variar y variará. Por ejemplo, la secuencia donante puede variar en longitud de varios nucleótidos a cientos de nucleótidos, a cientos de miles de nucleótidos.

20 Normalmente, la secuencia donante en el polinucleótido donante está flanqueada por una secuencia corriente arriba y una secuencia corriente abajo, que tienen identidad de secuencia sustancial con secuencias ubicadas corriente arriba y corriente abajo, respectivamente, de la secuencia que es el objetivo de la proteína de modificación de ADN programable. Debido a estas similitudes de secuencia, las secuencias corriente arriba y corriente abajo del polinucleótido donante permiten la recombinación homóloga entre el polinucleótido donante y la secuencia cromosómica que es el objetivo, de forma que la secuencia donante puede integrarse en (o intercambiarse con) la secuencia cromosómica.

25 La secuencia corriente arriba, como se usa en el presente documento, se refiere a una secuencia de ácido nucleico que comparte una identidad de secuencia sustancial con una secuencia cromosómica corriente arriba de la secuencia que es el objetivo de la proteína de modificación de ADN programable. De forma similar, secuencia corriente abajo se refiere a una secuencia de ácido nucleico que comparte una identidad de secuencia sustancial con una secuencia cromosómica corriente abajo de la secuencia que es el objetivo de la proteína de modificación de ADN programable. Como se usa en el presente documento, la frase "identidad de secuencia sustancial" se refiere a secuencias que tienen al menos aproximadamente el 75 % de identidad de secuencia. Por lo tanto, las secuencias corriente arriba y corriente abajo en el polinucleótido donante pueden tener aproximadamente el 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98% o 99% de identidad de secuencia con la secuencia corriente arriba o corriente abajo de la secuencia diana. En una realización ejemplar, las secuencias corriente arriba y corriente abajo en el polinucleótido donante pueden tener una identidad de secuencia de aproximadamente el 95 % o 100 % con secuencias cromosómicas corriente arriba o corriente abajo de la secuencia que es el objetivo de la proteína de modificación de ADN programable.

30 En algunas realizaciones, la secuencia corriente arriba comparte una identidad de secuencia sustancial con una secuencia cromosómica ubicada inmediatamente corriente arriba de la secuencia que es el objetivo de la proteína de modificación de ADN programable. En otras realizaciones, la secuencia corriente arriba comparte una identidad de secuencia sustancial con una secuencia cromosómica que está ubicada dentro de aproximadamente cien (100) nucleótidos corriente arriba de la secuencia diana. Por lo tanto, por ejemplo, la secuencia corriente arriba puede compartir una identidad de secuencia sustancial con una secuencia cromosómica que está ubicada de aproximadamente 1 a aproximadamente 20, de aproximadamente 21 a aproximadamente 40, de aproximadamente 41 a aproximadamente 60, de aproximadamente 61 a aproximadamente 80, o de aproximadamente 81 a aproximadamente 100 nucleótidos corriente arriba de la secuencia diana. En algunas realizaciones, la secuencia corriente abajo comparte una identidad de secuencia sustancial con una secuencia cromosómica ubicada inmediatamente corriente abajo de la secuencia que es el objetivo de la proteína de modificación de ADN programable. En otras realizaciones, la secuencia corriente abajo comparte una identidad de secuencia sustancial con una secuencia cromosómica que está ubicada dentro de aproximadamente cien (100) nucleótidos corriente abajo de la secuencia diana. Por lo tanto, por ejemplo, la secuencia corriente abajo puede compartir una identidad de secuencia sustancial con una secuencia cromosómica que está ubicada de aproximadamente 1 a aproximadamente 20, de aproximadamente 21 a aproximadamente 40, de aproximadamente 41 a aproximadamente 60, de aproximadamente 61 a aproximadamente 80, o de aproximadamente 81 a aproximadamente 100 nucleótidos corriente abajo de la secuencia diana.

35 Cada secuencia corriente arriba o corriente abajo puede variar en longitud de aproximadamente 20 nucleótidos a aproximadamente 5000 nucleótidos. En algunas realizaciones, las secuencias corriente arriba y corriente abajo pueden comprender aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800 o 5000 nucleótidos. En realizaciones específicas, las secuencias corriente arriba o corriente

abajo pueden variar en longitud de aproximadamente 50 a aproximadamente 1500 nucleótidos.

#### (e) Tipos de células

Son adecuadas para su uso en los procedimientos desvelados en el presente documento una diversidad de células. En general, la célula es una célula eucariota. Por ejemplo, la célula puede ser una célula de mamífero humano, una célula de mamífero no humano, una célula de vertebrado no mamífero, una célula de invertebrado, una célula de insecto, una célula vegetal, una célula de levadura o un organismo eucariota unicelular individual. En algunas realizaciones, la célula también puede ser un embrión unicelular. Por ejemplo, un embrión de mamífero no humano, incluyendo embriones de rata, hámster, roedor, conejo, felino, canino, ovino, porcino, bovino, equino y primate. En aún otras realizaciones, la célula puede ser una célula madre tal como las células madre embrionarias, células madre de tipo ES, células madre fetales, células madre de adulto, y similares. En una realización, la célula madre no es una célula madre embrionaria humana. Adicionalmente, las células madre pueden incluir las producidas por las técnicas desveladas en el documento WO2003/046141, o Chung y col. (Cell Stem Cell, 2008, 2:113-117). La célula puede estar *in vitro* o *in vivo* (es decir, dentro de un organismo). En realizaciones ejemplares, la célula es una célula de mamífero. En realizaciones particulares, la célula es una célula humana.

Los ejemplos no limitantes de células de mamífero adecuadas incluyen a las células embrionarias de riñón humano (HEK293, HEK293T); células de carcinoma de cuello uterino humano (HELA); células de pulmón humano (W138); células de hígado humano (Hep G2); células U2-OS de osteosarcoma humano, células A549 humanas, células A-431 humanas y células K562 humanas; células de ovario de hámster chino (CHO), células de riñón de cría de hámster (BHK); células NS0 de mieloma de ratón, células 3T3 de fibroblastos embrionarios de ratón (NIH3T3), células A20 de linfoma de linfocitos B de ratón; células B16 de melanoma de ratón; células C2C12 de mioblastos de ratón; células SP2/0 de mieloma de ratón; células C3H-10T1/2 mesenquimáticas embrionarias de ratón; células CT26 de carcinoma de ratón, células DuCuP de próstata de ratón; células EMT6 mamarias ratón; células Hepa1c1c7 de hepatoma de ratón; células J5582 de mieloma de ratón; células MTD-1A epiteliales de ratón; células MyEnd de miocardio del ratón; células RenCa renales de ratón; células RIN-5F pancreáticas de ratón; células de melanoma de ratón X64; células YAC-1 de linfoma de ratón; células 9L de glioblastoma de rata; células RBL de linfoma B de rata; células B35 de neuroblastoma de rata; células de hepatoma de rata (HTC); células de BRL 3A de hígado de rata búfalo; células de riñón canino (MDCK); células mamarias caninas (CMT); células D17 de osteosarcoma de rata; células DH82 de monocitos/macrófagos de rata; fibroblastos de riñón de mono transformados con SV-40 (COS7); células de riñón de mono CVI-76; células de riñón de mono verde africano (VERO-76). Se puede encontrar un extenso listado de líneas celulares de mamífero en el catálogo de la American Type Culture Collection (ATCC, Manassas, VA).

#### (IV) Aplicaciones

Las composiciones y procedimientos desvelados en el presente documento pueden usarse en una diversidad de aplicaciones terapéuticas, de diagnóstico, industriales y de investigación. En algunas realizaciones, la presente divulgación se puede usar para modificar cualquier secuencia cromosómica de interés en una célula, animal o planta, para modelar y/o estudiar la función de los genes, estudiar condiciones genéticas o epigenéticas de interés, o estudiar las rutas bioquímicas implicadas en diversas enfermedades o trastornos. Por ejemplo, se pueden crear organismos transgénicos que sean modelo de enfermedades o trastornos, en los que se altere la expresión de una o más secuencias de ácido nucleico asociadas con una enfermedad o trastorno. El modelo de enfermedad puede usarse para estudiar los efectos de las mutaciones en el organismo, estudiar el desarrollo y/o la progresión de la enfermedad, estudiar el efecto de un compuesto farmacéuticamente activo sobre la enfermedad y/o evaluar la eficacia de una estrategia potencial de terapia génica.

En otras realizaciones, las composiciones y los procedimientos se pueden usar para realizar estudios genómicos funcionales eficaces y rentables, que se pueden utilizar para estudiar la función de los genes implicados en un proceso biológico particular y cómo cualquier alteración en la expresión génica puede afectar el proceso biológico, o para realizar mutagénesis de exploración exhaustiva o por saturación de locus genómicos junto con un fenotipo celular. La mutagénesis de exploración exhaustiva o por saturación se puede utilizar para determinar características mínimas esenciales y vulnerabilidades discretas de elementos funcionales necesarios para la expresión génica, la resistencia a fármacos y la inversión de una enfermedad, por ejemplo.

En realizaciones adicionales, las composiciones y los procedimientos desvelados en el presente documento pueden usarse en pruebas de diagnóstico para establecer la presencia de una enfermedad o trastorno, y/o para su uso en la determinación de opciones de tratamiento. Los ejemplos de pruebas de diagnóstico adecuadas incluyen la detección de mutaciones específicas en células cancerosas (por ejemplo, mutación específica en EGFR, HER2 y similares), la detección de mutaciones específicas asociadas con enfermedades particulares (por ejemplo, repeticiones de trinucleótidos, mutaciones en la  $\beta$ -globina asociadas con la anemia de células falciformes, SNP específicos, etc.), detección de la hepatitis, detección de virus (por ejemplo., Zika), etc.

En realizaciones adicionales, las composiciones y procedimientos desvelados en el presente documento pueden usarse para corregir mutaciones genéticas asociadas con una enfermedad o trastorno particular, tal como, por ejemplo, corregir mutaciones del gen de la globina asociadas con la anemia de células falciformes o la talasemia, corregir mutaciones en el gen de la adenosina desaminasa asociadas con la inmunodeficiencia combinada grave (IDCG),

reducir la expresión de la HTT, el gen causante de la enfermedad de Huntington, o corregir mutaciones en el gen de la rodopsina para el tratamiento de la retinosis pigmentaria. Dichas modificaciones pueden hacerse en las células *ex vivo*.

- 5 En aún otras realizaciones, las composiciones y procedimientos desvelados en el presente documento pueden utilizarse para generar plantas de cultivo con rasgos mejorados o una resistencia aumentada al estrés ambiental. La presente divulgación también se puede utilizar para generar animales de granja con rasgos mejorados o animales de producción. Por ejemplo, los cerdos tienen muchas características que los hacen atractivos como modelos biomédicos, especialmente en medicina regenerativa o para el xenotrasplante.

## **DEFINICIONES**

- 10 A menos que se defina de otra manera, todos los términos técnicos y científicos usados en el presente documento tienen el significado que entiende comúnmente un experto en la materia a la que la presente invención pertenece. Las siguientes referencias proporcionan a los expertos una definición general de muchos de los términos utilizados en la presente invención: Singleton y col., Diccionario de Microbiología y Biología Molecular (2ª ed. 1994); El diccionario Cambridge de ciencia y tecnología (Ed. Walker, 1988); El glosario de genética, 5ª Ed., R. Rieger y col. (eds.), Springer Verlag (1991); y Hale y Marham, El diccionario Harper Collins de biología (1991). Como se usa en el presente documento, los siguientes términos tienen los significados que se les asigna a menos que se especifique otra cosa.

- 15 Cuando se introducen elementos de la presente divulgación o la realización (o realizaciones) preferente de la misma, los artículos "un", "una", "el", "la" y "dicho", "dicha" pretenden hacer referencia a uno o más de los elementos. Las expresiones "que comprende", "que incluye" y "que tiene" pretenden ser incluyentes y hacen referencia a que puede haber elementos adicionales además de los elementos enumerados.

20 El término "aproximadamente", cuando se usa en relación con un valor numérico,  $x$ , significa, por ejemplo,  $x \pm 5\%$ .

- 25 Como se usa en el presente documento, los términos "complementario" o "complementariedad" se refieren a la asociación de ácidos nucleicos bicatenarios mediante emparejamiento de bases a través de enlaces de hidrógeno específicos. El emparejamiento de bases puede ser emparejamiento de bases de Watson-Crick convencional (por ejemplo, pares 5'-A G T C-3' con la secuencia complementaria 3'-T C A G-5'). El emparejamiento de bases también puede ser por enlaces de hidrógeno de Hoogsteen o de Hoogsteen invertido. La complementariedad normalmente se mide con respecto a una región dúplex y, por lo tanto, por ejemplo, excluye protuberancias. La complementariedad entre dos cadenas de la región dúplex puede ser parcial y expresarse como un porcentaje (por ejemplo, del 70 %), si solo algunas (por ejemplo, el 70 %) de las bases son complementarias. Las bases que no son complementarias están "desapareadas". La complementariedad también puede ser completa (es decir, del 100 %), si todas las bases en la región dúplex son complementarias.

30 Como se usa en el presente documento, la expresión "sistema CRISPR/Cas" se refiere a un complejo que comprende una proteína CRISPR/Cas (es decir, una nucleasa, nickase o una proteína catalíticamente muerta) y un ARN de guía.

- 35 La expresión "secuencia endógena", como se usa en el presente documento, se refiere a una secuencia cromosómica que es nativa de la célula.

Como se usa en el presente documento, el término "exógeno" se refiere a una secuencia que no es nativa de la célula, o a una secuencia cromosómica cuya ubicación nativa en el genoma de la célula está en una ubicación cromosómica distinta.

- 40 Un "gen", como se usa en el presente documento, se refiere a una región de ADN (incluidos exones e intrones) que codifica un producto génico, así como todas las regiones de ADN que regulan la producción del producto génico, ya sea que tales secuencias reguladoras sean o no adyacentes a las secuencias codificadas y/o transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción, tales como sitios de unión al ribosoma y sitios internos de entrada al ribosoma, potenciadores, silenciadores, aislantes, elementos limítrofes, orígenes de replicación, sitios de unión de matriz y regiones de control de locus.

45 El término "heterólogo" se refiere a una entidad que no es endógena ni nativa en la célula de interés. Por ejemplo, una proteína heteróloga se refiere a una proteína que se obtiene o se obtuvo originalmente de una fuente exógena, tal como una secuencia de ácido nucleico introducida de forma exógena. En algunos casos, la proteína heteróloga normalmente no es producida por la célula de interés.

- 50 Las expresiones "estructura local de la cromatina" o "configuración local de la cromatina", como se usa en el presente documento, se refiere a la estructura de los nucleosomas y/o al espaciamiento de las proteínas histona, y generalmente no se refiere a la compactación de los nucleosomas en fibras de cromatina y heterocromatina.

- 55 El término "nickasa" se refiere a una enzima que escinde una cadena de una secuencia de ácido nucleico bicatenario (es decir, mella una secuencia bicatenaria). Por ejemplo, una nucleasa con actividad de escisión bicatenaria puede modificarse por mutación y/o delección para que funcione como una nickasa y escinda solo una cadena de una

secuencia bicatenaria.

El término "nucleasa", como se usa en el presente documento, se refiere a una enzima que escinde ambas cadenas de una secuencia de ácido nucleico bicatenario.

5 Las expresiones "ácido nucleico" y "polinucleótido" se refieren a un polímero de desoxirribonucleótidos o de ribonucleótidos, con una conformación lineal o circular, y en forma monocatenaria o bicatenaria. Para los fines de la presente divulgación, estos términos no deben considerarse como limitantes con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que se modifican en las fracciones de la base, azúcar y/o fosfato (por ejemplo, esqueletos de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de emparejamiento de bases; es decir, un análogo de A tendrá un emparejamiento de bases con T.

15 El término "nucleótido" se refiere a desoxirribonucleótidos o a ribonucleótidos. Los nucleótidos pueden ser nucleótidos convencionales (es decir, adenosina, guanosina, citidina, timidina y uridina), isómeros de nucleótidos o análogos de nucleótidos. Un análogo de nucleótido se refiere a un nucleótido que tiene una base purina o pirimidina modificada o una fracción de ribosa modificada. Un análogo de nucleótido puede ser un nucleótido de origen natural (por ejemplo, inosina, pseudouridina, etc.) o un nucleótido que no sea de origen natural. Los ejemplos no limitantes de modificaciones en las fracciones de azúcar o base de un nucleótido incluyen la adición (o eliminación) de grupos acetilo, grupos amino, grupos carboxilo, grupos carboximetilo, grupos hidroxilo, grupos metilo, grupos fosforilo y grupos tiol, así como la sustitución de los átomos de carbono y nitrógeno de las bases por otros átomos (por ejemplo, 7-desaza purinas). Los análogos de nucleótidos también incluyen didesoxinucleótidos, 2'-O-metil nucleótidos, ácidos nucleicos bloqueados (LNA), ácidos péptidonucleicos (PNA) y morfolinós.

20 Los términos "polipéptido" y "proteína" se usan de forma indistinta para hacer referencia a un polímero de restos de aminoácido.

25 La expresión "sitio próximo", como se usa en el presente documento, se refiere a un sitio de unión o secuencia de nucleótidos que está ubicado dentro de aproximadamente 250 pares de bases a cada lado de una secuencia diana en el ADN cromosómico.

Como se usa en el presente documento, la expresión "proteína de modificación de ADN programable" se refiere a una proteína que se genomanipula para que se una a una secuencia diana específica en el ADN cromosómico y que modifica el ADN o una proteína (o proteínas) asociada con el ADN en o cerca de la secuencia diana.

30 La expresión "proteína de unión a ADN programable" como se usa en el presente documento, se refiere a una proteína que está genomanipulada para unirse a una secuencia diana específica en el ADN cromosómico, pero no modifica el ADN o la proteína (o proteínas) asociada con el ADN en o cerca de la secuencia diana.

35 Las expresiones "secuencia diana", "secuencia cromosómica diana", y "sitio diana", se usan indistintamente para referirse a la secuencia específica en el ADN cromosómico a la que se dirige la proteína de modificación de ADN programable, y el sitio en el que la proteína de modificación de ADN programable modifica el ADN o la proteína (o proteínas) asociada con el ADN.

40 Las técnicas para determinar la secuencia del ácido nucleico y de aminoácidos son conocidas en la técnica. Normalmente, tales técnicas incluyen determinar la secuencia de nucleótidos del ARNm para un gen y/o determinar la secuencia de aminoácidos codificada por el mismo, y comparar estas secuencias con una segunda secuencia de nucleótidos o aminoácidos. Las secuencias genómicas también se pueden determinar y comparar de esta manera. En general, identidad se refiere a una correspondencia exacta de nucleótido a nucleótido o de aminoácido a aminoácido de dos secuencias polinucleótidos o de polipéptidos, respectivamente. Se pueden comparar dos o más secuencias (polinucleótídicas o de aminoácidos) determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sea secuencias de ácido nucleico o de aminoácidos, es el número de coincidencias exactas entre dos secuencias alineadas dividido por la longitud de las secuencias más cortas y multiplicado por 100. El algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2:482-489 (1981), proporciona un alineamiento aproximado para secuencias de ácido nucleico. Este algoritmo se puede aplicar a secuencias de aminoácidos utilizando la matriz de puntuación desarrollada por Dayhoff, *Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 supl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., Estados Unidos, y normalizado por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). Una implementación ejemplar de este algoritmo para determinar el porcentaje de identidad de una secuencia la proporciona Genetics Computer Group (Madison, Wis.) en la aplicación de utilidad "BestFit". Otros programas adecuados para calcular el porcentaje de identidad o similitud entre secuencias son generalmente conocidos en la técnica, por ejemplo, otro programa de alineamiento es BLAST, utilizado con parámetros predeterminados. Por ejemplo, BLASTN y BLASTP se pueden usar con los siguientes parámetros predeterminados: código genético = estándar; filtro=ninguno; cadenas=ambas; corte=60; expectativa=10; Matriz= BLOSUM62; Descripciones=50 secuencias; clasificación por= PUNTUACIÓN ALTA; Bases de datos=no redundantes, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Los detalles de estos programas se pueden encontrar en el sitio web de GenBank.

Como se pueden realizar varios cambios en las células y procedimientos descritos anteriormente sin apartarse del

ámbito de la invención, se pretende que toda la materia contenida en la descripción anterior y en los ejemplos que se dan a continuación, se interprete como ilustrativa y no en un sentido limitante.

### **Ejemplos**

Los siguientes ejemplos ilustran determinados aspectos de la divulgación.

#### **5 Ejemplo 1. Potenciación de la edición génica de CRISPR-Cas9 de *Francisella novicida* (FnCas9)**

FnCas9 es una CRISPR-Cas9 de tipo IIB. Presenta una especificidad intrínseca más alta que la SpCas9 ampliamente utilizada, pero se ha descubierto que en células humanas es menos consistente que SpCas9. Para determinar si la unión de proteínas de unión a ADN programables a sitios próximos podría permitir a la nucleasa escindir una diana inaccesible (es decir, el locus POR) en células humanas, se transfectaron células K562 con 5,6 µg de ADN plasmídico de FnCas9, 5 µg de ADN plasmídico de SpCas9 catalíticamente muerta (SpdCas9) y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células (véase la FIG. 2). Se recogió el ADN genómico 3 días después de la transfección y la región diana se amplificó por PCR con el cebador directo 5'-CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:9) y el cebador inverso 5'-ACAGGTCGTGGACTCACA-3' (SEQ ID NO:10). Las inserciones/deleciones (indeles) dirigidas por FnCas9 se determinaron mediante digestión con nucleasa Cel-I y análisis en gel de poliacrilamida.

15 Como se muestra en la FIG. 2, FnCas9 no pudo escindir la diana cuando se transfectó sola. No obstante, cuando se transfectó en combinación con SpdCas9 para ayudar a alterar la configuración local de la cromatina, FnCas9 pudo escindir la diana a niveles consistentes, con un 10-11 % de indeles, cuando se usó SpdCas9 para unirse a un sitio próximo. Cuando se usó SpdCas9 para unirse dos sitios próximos, la actividad de FnCas9 aumentó aún más, al 28 % de indeles. Estos resultados demuestran que el procedimiento desvelado en el presente documento puede permitir  
20 que una endonucleasa escinda de manera eficaz una diana que de otro modo sería inaccesible, y que existe un efecto sinérgico entre los dos sitios utilizados para alterar la configuración local de cromatina local.

#### **Ejemplo 2. Potenciación de la edición génica de CRISPR-Cas9 de *Campylobacter jejuni* (CjCas9)**

CjCas9 es una CRISPR-Cas9 de tipo IIC. Es la Cas9 más pequeña caracterizada hasta ahora y tiene un único requisito de PAM ACAY. No obstante, se ha descubierto que la nucleasa es inactiva en la mayoría de dianas en las células humanas. Para determinar si los procedimientos desvelados en el presente documento podrían permitir que la proteína CjCas9 se una a una diana inaccesible en células humanas, se transfectaron células K562 con 4,2 µg de ADN plasmídico de CjCas9 catalíticamente muerta (CjdCas9) etiquetada con Flag, 5 µg de ADN plasmídico de SpCas9 catalíticamente muerta (SpdCas9) y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células (véase la FIG. 3A). Las células se fijaron en formaldehído 16 horas después de la transfección y se llevó a cabo una inmunoprecipitación de cromatina (ChIP, forma siglada de *chromatin immunoprecipitation*, inmunoprecipitación de cromatina) utilizando un anticuerpo anti-flag. La unión a la diana mediante Flag-CjdCas9 se determinó mediante PCR digital de gotitas (PCRdd).

35 Como se muestra en la FIG. 3C, Flag-CjdCas9 pudo unirse a una diana accesible previamente conocida en el locus AAVS1, pero no pudo unirse a una diana inaccesible en el locus POR cuando se transfectó sola. Sin embargo, cuando se transfectó en combinación con SpdCas9 para alterar la configuración local de la cromatina, Flag-CjdCas9 pudo unirse a la diana en POR de forma incluso más eficaz que su unión a la diana en AAVS1.

Para examinar el efecto sobre la escisión del ADN diana, se transfectaron células K562 con 4,2 µg de ADN plasmídico de CjCas9, 5 µg de ADN plasmídico de SpdCas9 y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células. Se recogió el ADN genómico 3 días después de la transfección y la región diana se amplificó por PCR con el cebador directo 5'-CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:9) y el cebador inverso 5'-ACAGGTCGTGGACTCACA-3' (SEQ ID NO:10). La actividad de escisión de CjCas9 en la diana POR se determinó mediante digestión con nucleasa Cel-I y análisis en gel de poliacrilamida. Como se muestra en la FIG. 4, CjCas9 no pudo escindir la diana sin SpdCas9. No obstante, cuando se transfectó en combinación con SpdCas9, CjCas9 pudo escindir la diana de forma eficaz con un 34,1-37,9 % de indeles. Estos resultados demuestran que el procedimiento desvelado en el presente documento puede permitir que una nucleasa se una y escinda de forma eficaz una diana que de otro modo sería inaccesible.

#### **45 Ejemplo 3. Potenciación de la edición génica de Cpf1 de *Francisella novicida* (FnCpf1)**

FnCpf1 es un sistema CRISPR-Cas de tipo V. Los sistemas Cpf1 son significativamente divergentes de los sistemas CRISPR-Cas9 de tipo II. A diferencia de los sistemas Cas9, los sistemas Cpf1 utilizan un PAM rico en T 5' y un ARN de guía único para el direccionamiento, sin un ARN<sub>tracr</sub> (Zetsche y col., Cell, 2015, 163:1-13). Estos sistemas CRISPR "más nuevos" tienen el potencial de simplificar aún más la práctica de la edición de genes, pero se ha descubierto que muchos sistemas Cpf1 están inactivos en las células humanas. Para determinar si los procedimientos desvelados en el presente documento podrían permitir que la una nucleasa divergente "inactiva" Cpf1 escinda dianas endógenas en células humanas, se transfectaron células K562 con 5 µg de ADN plasmídico de FnCpf1, 5 µg de ADN plasmídico de SpdCas9 y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células (véase la FIG. 5). Se recogió el ADN genómico 3 días después de la transfección y la región diana se amplificó por PCR con el cebador directo 5'-CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:9) y el cebador inverso 5'-ACAGGTCGTGGACTCACA-3' (SEQ ID NO:10). La actividad de escisión de FnCpf1 en una diana POR se determinó mediante digestión con nucleasa Cel-I y análisis en gel de poliacrilamida.

Como se muestra en la FIG. 5, FnCpf1 no pudo escindir la diana cuando se transfectó sola, pero pudo escindir la diana de forma eficaz cuando se transfectó en combinación con SpdCas9. Estos resultados demuestran que el procedimiento desvelado en el presente documento es aplicable a sistemas CRISPR-Cas de tipo V divergentes.

#### **Ejemplo 4. Edición selectiva entre dianas idénticas en HBB y HBD humanos**

5 Se utilizaron dos dianas idénticas en humano (es decir, HBB y HBD) para determinar si los procedimientos desvelados en el presente documento podrían facilitar la edición selectiva entre sitios idénticos en genes distintos. Se transfectaron células K562 con 4,2 µg de ADN plasmídico de CjCas9, 5 µg de ADN plasmídico de SpdCas9 y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células (véase la FIG. 6). Se recogió el ADN genómico 3 días después de la transfección y las dos regiones diana se amplificaron por PCR con el cebador directo 5'-CGGCTGTCATCACTTAGACCTCA-3' (SEQ ID NO:11) y el cebador inverso 5'-GCAGCCTAAGGGTGGGAAAATAGA-3' (SEQ ID NO:12) para HBB, y el cebador directo 5'-AGGGCAAGTTAAGGGAATAGTGGAA-3' (SEQ ID NO:13) y el cebador inverso 5'-CCAAGGGTAGACCACAGTAATCTG-3' (SEQ ID NO:14) para HBD. La actividad de escisión de CjCas9 en las dianas de HBB y HBD se determinó mediante digestión con nucleasa Cel-I y análisis en gel de poliacrilamida.

15 Como se muestra en la FIG. 6, cuando se transfectó sola, CjCas9 no pudo escindir ninguna de las dianas. No obstante, cuando se transfectó en combinación con SpdCas9 dirigida a sitios próximos a HBB, CjCas9 escindió la diana objetivo HBB de manera eficaz, pero no pudo escindir la diana HBD idéntica. Los SNP presentes en la población de células K562 provocaron las dos bandas de digestión con nucleasa Cel-I en los dos primeros carriles. Estos resultados demuestran la capacidad exclusiva del procedimiento desvelado para mejorar la selectividad de edición de genes.

#### **Ejemplo 5. Potenciación de la edición génica de CRISPR-Cas9 de *Streptococcus pyogenes* (SpCas9)**

SpCas9 es una CRISPR-Cas9 de tipo IIA, y se ha utilizado extensamente en la modificación del genoma debido a su consistente actividad en células eucariotas. Sin embargo, su actividad también puede variar ampliamente de una diana a otra. Para determinar si los procedimientos desvelados en el presente documento también podrían potenciar a esta nucleasa, se transfectaron células K562 con 5 µg de ADN plasmídico de SpCas9, 5,6 µg de FnCas9 catalíticamente muerta (FnCas9) y 3 µg de ADN plasmídico de cada ARN<sub>g</sub> por millón de células (véase la FIG. 7). Se recogió el ADN genómico 3 días después de la transfección y la región diana se amplificó por PCR con el cebador directo 5'-CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:9) y el cebador inverso 5'-ACAGGTCGTGGACTCACA-3' (SEQ ID NO:10). La actividad de escisión de SpCas9 en la diana POR se determinó mediante digestión con nucleasa Cel-I y análisis en gel de poliacrilamida.

30 Como se muestra en la FIG. 7, La actividad de escisión de SpCas9 aumentó significativamente cuando se transfectó en combinación con FnCas9, en comparación con cuando se transfectó sola. Estos resultados muestran que el procedimiento desvelado en el presente documento también puede aplicarse a endonucleasas robustas.

#### **Ejemplo 6. Potenciación de la edición de genes utilizando un donante de oligonucleótido de ADN<sub>mc</sub>**

35 Se transfectaron células K562 con 4,2 µg de ADN plasmídico de CjCas9, 5 µg de ADN plasmídico de SpdCas9, 3 µg de ADN plasmídico de cada ARN<sub>g</sub> y 300 pmol de un donante de oligonucleótido de ADN<sub>mc</sub> de 88 nt para la integración dirigida de un sitio de restricción *EcoRI* por millón de células. Se recogió el ADN genómico 3 días después de la transfección y la región diana se amplificó por PCR con el cebador directo 5'-CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:9) y el cebador inverso 5'-ACAGGTCGTGGACTCACA-3' (SEQ ID NO:10). La integración dirigida del sitio de restricción *EcoRI* se determinó por digestión con la enzima de restricción *EcoRI* y análisis por gel de poliacrilamida. Como se muestra en la FIG. 8, el sitio de restricción se integró de manera eficaz (28-37 %) en el locus POR cuando el donante de oligonucleótido de ADN<sub>mc</sub> se transfectó en concierto con CjCas9 y SpdCas9, mientras que no se detectó integración cuando el donante de oligonucleótido se transfectó solo o en combinación con CjCas9 sin SpdCas9. Estos resultados demuestran que el procedimiento desvelado en el presente documento puede facilitar una edición de genes eficaz utilizando un donante de oligonucleótido de ADN<sub>mc</sub> sobre una diana que de otro modo sería inaccesible.

#### **Ejemplo 7. Potenciación de la detección de ADN genómico específica de secuencia en células vivas y fijadas**

La fusión de proteínas Cas9 a proteínas fluorescentes ha permitido la detección de la dinámica cromosómica en células vivas (Chen y col., Cell, 2013, 155:1479-91). Por lo tanto, se cree que la dinámica estructural de la cromatina influirá sobre la capacidad de los complejos del sistema CRISPR/Cas para acceder a diversos locus genómicos. Por lo tanto, se cree que la colocación de complejos de CRISPR (dCas9) próximos a los que albergan dCas9-GFP potencia la detección de la dinámica cromosómica en un grado similar a lo observado en el Ejemplo 2 para la inmunoprecipitación de cromatina. Por ejemplo, CjdCas9 puede fusionarse con GFP y dirigirse a una región con un estado de la cromatina que impida la unión detectable de CjdCas9-GFP. El sistema basado en SpdCas9 se puede proyectar cerca de las dianas de CjdCas9-GFP para producir una señal detectable. Para regiones de cromatina que son resistentes a la unión y detección de SpdCas9-GFP, se puede usar una molécula próxima de FnCas9 para potenciar la detección, en un grado similar al mostrado en el Ejemplo 5 para el direccionamiento próximo de SpCas9 y FnCas9 y la potenciación de la actividad de rotura bicatenaria. Adicionalmente, dado que estudios previos han indicado que los requisitos sobre el grado de hibridación entre el ARN de guía de CRISPR y el ADN genómico puede ser menor para la unión que para

la escisión bicatenaria (Wu y col., Nature Biotechnology, 2014, 32(7): 670-6), se cree que el uso de la unión próxima de CRISPR aumenta las relaciones señal-ruido para la detección de ADN genómico en las células.

5 Se han aplicado a células fijadas procedimientos de detección similares basados en CRISPR (Deng y col., Proc. Natl. Acad. Sci. USA, 2015, 112(38):11870-75). Por lo tanto, se cree que el direccionamiento proximal de CRISPR potenciará la detección de ADN fijado de una manera similar a la descrita anteriormente para las células vivas. Dado que las cadenas de ADN genómico en las células fijadas están reticuladas de forma química, la interrogación de la información de secuencias por hibridación de sondas de ácido nucleico normalmente precisa una etapa de pretratamiento térmico o procesamiento químico para separar las hebras lo suficiente. Por lo tanto, es posible que el direccionamiento proximal de CRISPR haga que el ADN fijado sea más accesible y se reduzca la grado (o requisito) de tratamiento térmico o químico de las células fijadas. La eliminación del tratamiento térmico o químico proporcionaría ventajas en la simplificación del protocolo de diagnóstico y el mantenimiento de las estructuras moleculares intracelulares que reflejan mejor la biología de las células vivas y, por lo tanto, resultados de diagnóstico más fundados.

**Ejemplo 8. Potenciación de la activación y represión de genes basadas en CRISPR en células eucariotas**

15 La fusión de proteínas Cas9 a dominios de regulación de la transcripción ha permitido la activación y represión de genes específicos (Konermann y col., Nature, 2014; 517(7536):583-8; Gilbert y col., Cell, 2014, 159(3):547-661). Se cree que la dinámica estructural de la cromatina influirá sobre la capacidad del complejo de CRISPR para acceder a diversos locus genómicos e inducir la activación o la represión. Por lo tanto, se cree que la colocación de complejos de CRISPR (dCas9) próximos a los que albergan dCas9 fusionada a dominios de regulación de la transcripción potencia la regulación dirigida de genes en un grado similar a lo observado en el Ejemplo 2 para la inmunoprecipitación de cromatina. Para regiones de cromatina que son resistentes a la unión y la modificación mediante SpdCas9-reguladores de la transcripción, se puede usar una molécula próxima de FndCas9 para potenciar la activación o represión génica en un grado similar al mostrado en el Ejemplo 5 para el direccionamiento próximo de SpCas9 y FndCas9 y la potenciación de la actividad de rotura bicatenaria.

**Ejemplo 9. Potenciación de la modificación epigenética basada en CRISPR en células eucariotas**

25 La fusión de las proteínas Cas9 a los dominios de modificación epigenética ha permitido modificaciones cromosómicas epigenéticas dirigidas, tales como la acetilación de histonas por p300 o la desaminación de citosina por la citosina desaminasa (Hilton y col., Nat. Biotechnol; 2015, 33(5):510-7; Komor y col., Nature, 2016, 533(7603):420-4). Se cree que la dinámica estructural de la cromatina influirá sobre la capacidad de los complejos de CRISPR para acceder a diversos locus genómicos. Por lo tanto, la colocación de complejos de CRISPR (dCas9) próximos a los que albergan dCas9 fusionada a modificadores epigenéticos debería potenciar la modificación epigenética dirigida del ADN cromosómico, proteínas locales o ARN local en un grado similar a lo observado en el Ejemplo 2 para inmunoprecipitación de cromatina. Para regiones de cromatina que son resistentes a la unión y la modificación mediante SpdCas9-modificadores epi., se puede usar una molécula próxima de FndCas9 para potenciar la detección en un grado similar al mostrado en el Ejemplo 5 para el direccionamiento próximo de SpCas9 y FndCas9 y la potenciación de la actividad de rotura bicatenaria.

LISTADO DE SECUENCIAS

<110> SIGMA-ALDRICH CO. LLC

<120> USO DE PROTEÍNAS DE UNIÓN A ADN PROGRAMABLES PARA POTENCIAR LA MODIFICACIÓN DIRIGIDA DEL GENOMA

40 <130> P2970EP00

<150> US 62/358.415

<151> 05/07/2016

<150> US 62/344.858

<151> 02/06/2016

45 <160> 49

<170> PatentIn versión 3.5

<210> 1

<211> 7

<212> PRT

50 <213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 1

ES 2 760 477 T3

Pro Lys Lys Lys Arg Lys Val  
1 5

5 <210> 2  
<211> 7  
<212> PRT  
<213> Secuencia artificial

<220>  
<223> SINTETIZADO

<400> 2

Pro Lys Lys Lys Arg Arg Val  
1 5

10 <210> 3  
<211> 16  
<212> PRT  
<213> Secuencia artificial

15 <220>  
<223> SINTETIZADO

<400> 3

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys  
1 5 10 15

20 <210> 4  
<211> 20  
<212> PRT  
<213> Secuencia artificial

<220>  
<223> SINTETIZADO

<400> 4

25 Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys  
1 5 10 15

Lys Arg Lys Val 20

30 <210> 5  
<211> 19  
<212> PRT  
<213> Secuencia artificial

<220>  
<223> SINTETIZADO

<400> 5

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys  
1 5 10 15

Arg Lys Val

35 <210> 6  
<211> 24  
<212> PRT  
<213> Secuencia artificial

40 <220>  
<223> SINTETIZADO

<400> 6

ES 2 760 477 T3

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly  
1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val  
20

5 <210> 7  
<211> 27  
<212> PRT  
<213> Secuencia artificial

<220>  
<223> SINTETIZADO

<400> 7

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly  
1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val  
20 25

10 <210> 8  
<211> 21  
<212> PRT  
<213> Secuencia artificial

15 <220>  
<223> SINTETIZADO

<400> 8

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys  
1 5 10 15

Lys Lys Arg Lys Val  
20

20 <210> 9  
<211> 21  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> SINTETIZADO

<400> 9

25 ctcccctgct tctgtcgta t 21

<210> 10  
<211> 20  
<212> ADN  
<213> Secuencia artificial

30 <220>  
<223> SINTETIZADO

<400> 10

acaggctgtg gacactcaca20

<210> 11

ES 2 760 477 T3

<211> 23  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 5 <223> SINTETIZADO  
  
 <400> 11  
 cggctgcat cacttagacc tca 23  
  
 <210> 12  
 10 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 <400> 12  
 15 gcagcctaag ggtgggaaaa taga 24  
  
 <210> 13  
 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
  
 20 <220>  
 <223> SINTETIZADO  
  
 <400> 13  
 agggcaagtt aaggaatag tggaa 25  
  
 <210> 14  
 25 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 30 <400> 14  
 ccaaggtag accaccagta atctg 25  
  
 <210> 15  
 <211> 101  
 35 <212> ADN  
 <213> *Homo sapiens*  
  
 <400> 15  
 cctttccagc attcgccagt acgagcttgt ggtccacacc gacatagatg cggccaaggt 60  
 gtacatgggg gagatgggcc ggctgaagag ctacgagaac c 101  
  
 <210> 16  
 <211> 101  
 40 <212> ADN  
 <213> *Homo sapiens*  
  
 <400> 16  
 ggttctcgta gctcttcagc cggcccatct ccccatgta caccttgGCC gcatctatgt 60  
 cgggtgtggac cacaagctcg tactggcgaa tgctggaaag g 101  
  
 <210> 17

ES 2 760 477 T3

<211> 105  
 <212> ADN  
 <213> *Homo sapiens*  
 <400> 17  
 ccgcttcccc gcctcaccct tggctctccc tttccagcat tcgccagtac gagcttgtgg 60  
 5 tccacaccga catagatgcg gccaaagggtg acatggggga gatgg 105  
 <210> 18  
 <211> 105  
 <212> ADN  
 <213> *Homo sapiens*  
 10 <400> 18  
 ccatctcccc catgtacacc ttggccgcat ctatgtcggg gtggaccaca agctcgtact 60  
 ggcgaatgct ggaaagggga gaccaagggt gaggccggga agcgg 105  
 <210> 19  
 <211> 110  
 <212> ADN  
 <213> *Homo sapiens*  
 15 <400> 19  
 ccgcttcccc gcctcaccct tggctctccc tttccagcat tcgccagtac gagcttgtgg 60  
 tccacaccga catagatgcg gccaaagggtg acatggggga gatgggcccgg 110  
 <210> 20  
 <211> 110  
 <212> ADN  
 <213> *Homo sapiens*  
 20 <400> 20  
 ccggcccatc tccccatgt acaccttggc cgcactatg tcggtgtgga ccacaagctc 60  
 gtactggcga atgctggaaa ggggagacca aggggtgaggc cgggaagcgg 110  
 <210> 21  
 <211> 117  
 <212> ADN  
 <213> *Homo sapiens*  
 25 <400> 21  
 tcgactgttg cttacacttt cttctgacat aacagtgttc actagcaacc tcaaacagac 60  
 accatgggtgc atctgactcc tgaggagaag actgctgtca atgccctgtg gggcaaa 117  
 30 <210> 22  
 <211> 117  
 <212> ADN  
 <213> *Homo sapiens*  
 <400> 22  
 gaacgggggtg tcccgtcatt gccgtctgaa gaggagtcc cagtctacgt ggtaccacag 60  
 35 acaaactcca acgatcactt gtgtcaacac agtcttcggt tacattcgtt atctacc 117  
 <210> 23  
 <211> 142

ES 2 760 477 T3

<212> ADN  
 <213> *Homo sapiens*  
  
 <220>  
 <221> misc\_feature  
 <222> (74)..(113)  
 <223> n es a, c, g o t  
  
 <400> 23  
  
 tgtacatggg ggagatgggc cggctgaaga gctacgagaa ccagaagccg tgagtggagg 60  
 gagcgtggct tggnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnntctgccg 120  
 tgtatcccca tatccccaca gg 142  
  
 <210> 24  
 <211> 142  
 <212> ADN  
 <213> *Homo sapiens*  
  
 <220>  
 <221> misc\_feature  
 <222> (30)..(69)  
 <223> n es a, c, g o t  
  
 <400> 24  
  
 cctgtgggga tatggggata cacggcagan nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60  
 nnnnnnnnnc caagccaacgc tccctccact cacggcttct ggttctcgta gctcttcagc 120  
 cggcccatct ccccatgta ca 142  
  
 <210> 25  
 <211> 47  
 <212> ADN  
 <213> *Homo sapiens*  
  
 <400> 25  
  
 caggatcct gtgtcccga gctggacca ccttatatc ccagggc 47  
  
 <210> 26  
 <211> 47  
 <212> ADN  
 <213> *Homo sapiens*  
  
 <400> 26  
  
 gccctgcaa tataaggtg tccagctcg gggacacagg atccctg 47  
  
 <210> 27  
 <211> 88  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 <400> 27  
  
 cacccttggg ctccccttc cagcattcgc cagtacgagc gaattcttgt ggtccacacc 60  
  
 gacatagatg cggccaaggt gtacatgg 88

ES 2 760 477 T3

<210> 28  
 <211> 11  
 <212> PRT  
 <213> Secuencia artificial  
 5 <220>  
 <223> SINTETIZADO  
 <400> 28  
  
                   Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg  
                   1                                  5                                  10  
  
 10 <210> 29  
 <211> 9  
 <212> PRT  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 15 <400> 29  
  
                   Arg Lys Lys Arg Arg Gln Arg Arg Arg  
                   1                                  5  
  
 20 <210> 30  
 <211> 9  
 <212> PRT  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 <400> 30  
  
                   Pro Ala Ala Lys Arg Val Lys Leu Asp  
                   1                                  5  
  
 25 <210> 31  
 <211> 11  
 <212> PRT  
 <213> Secuencia artificial  
  
 <220>  
 30 <223> SINTETIZADO  
  
 <400> 31  
  
                   Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro  
                   1                                  5                                  10  
  
 35 <210> 32  
 <211> 8  
 <212> PRT  
 <213> Secuencia artificial  
  
 <220>  
 <223> SINTETIZADO  
  
 <400> 32  
  
                   Val Ser Arg Lys Arg Pro Arg Pro  
                   1                                  5  
  
 40 <210> 33  
 <211> 8



ES 2 760 477 T3

<220>  
 <223> SINTETIZADO  
 <400> 38

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg  
 1 5 10

5 <210> 39  
 <211> 20  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO  
 <400> 39

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys  
 1 5 10 15

Lys Ser Lys Lys  
 20

15 <210> 40  
 <211> 17  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO  
 <400> 40

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys  
 1 5 10 15

Lys

25 <210> 41  
 <211> 38  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO  
 <400> 41

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly  
 1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro  
 20 25 30

Arg Asn Gln Gly Gly Tyr  
 35

30 <210> 42  
 <211> 42  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO  
 <400> 42

35 <210> 43  
 <211> 40  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO  
 <400> 43

ES 2 760 477 T3

<400> 42

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu  
 1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys  
 20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val  
 35 40

5

<210> 43  
 <211> 11  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO

<400> 43

10

Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala  
 1 5 10

<210> 44  
 <211> 11  
 <212> PRT  
 <213> Secuencia artificial

15

<220>  
 <223> SINTETIZADO

<400> 44

Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg Arg  
 1 5 10

20

<210> 45  
 <211> 11  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> SINTETIZADO

25

<400> 45

Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg Arg  
 1 5 10

<210> 46  
 <211> 12  
 <212> PRT  
 <213> Secuencia artificial

30

<220>  
 <223> SINTETIZADO

<400> 46

Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg  
 1 5 10

35

<210> 47  
 <211> 27  
 <212> PRT  
 <213> Secuencia artificial

ES 2 760 477 T3

<220>

<223> SINTETIZADO

<400> 47

Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu  
1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu  
20 25

5

<210> 48

<211> 33

<212> PRT

<213> Secuencia artificial

<220>

10

<223> SINTETIZADO

<400> 48

Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala  
1 5 10 15

Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu  
20 25 30

Ala

<210> 49

<211> 16

15

<212> PRT

<213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 49

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys  
1 5 10 15

20

## REIVINDICACIONES

1. Una composición que comprende:

(a) un sistema de nucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) guiado por ARN o un ácido nucleico que codifica dicho sistema de nucleasa CRISPR, en el que el sistema de nucleasa CRISPR comprende (i) una proteína de modificación de ADN programable que es una proteína CRISPR, y (ii) un ARN de guía; y

(b) al menos un sistema CRISPR catalíticamente inactivo o un ácido nucleico que codifica dicho al menos un sistema CRISPR catalíticamente inactivo, en el que cada sistema CRISPR catalíticamente inactivo comprende (i) una proteína de unión a ADN programable que es una proteína CRISPR catalíticamente inactiva y (ii) un ARN de guía; y en la que

(x) la proteína de modificación de ADN programable es una proteína CRISPR de tipo II y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II, o

(y) la proteína de modificación de ADN programable es una proteína CRISPR de tipo V y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II; y en la que

la composición como se define anteriormente no se desvela en los documentos WO2017/070598 o WO2017/096328.

2. La composición de acuerdo con la reivindicación 1, en la que (i) la proteína CRISPR de tipo II se selecciona de CRISPR-Cas9 de *Francisella novicida* (FnCas9), CRISPR-Cas9 de *Campylobacter jejuni* (CjCas9) y CRISPR-Cas9 de *Streptococcus pyogenes* (SpCas9); y/o (ii) la proteína CRISPR de tipo V es CRISPR-Cpf1 de *Francisella novicida* (FnCpf1).

3. La composición de acuerdo con la reivindicación 1 o la reivindicación 2, en la que la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II Cas9, en la que la proteína Cas9 tiene una o más mutaciones en cada uno del dominio de tipo RuvC y el dominio de tipo HNH.

4. La composición de acuerdo con la reivindicación 3, en la que (i) la una o más mutaciones en el dominio de tipo RuvC son D10A, D8A, E762A y/o D986A; y/o (ii) la una o más mutaciones en el dominio de tipo HNH son H840A, H559A, N854A, N856A y/o N863A.

5. La composición de cualquier reivindicación precedente, en la que el ácido nucleico que codifica cada proteína CRISPR es ARNm o ADN.

6. La composición de cualquier reivindicación precedente, en la que el ácido nucleico que codifica cada proteína CRISPR es parte de un vector plasmídico o un vector vírico.

7. La composición de cualquier reivindicación precedente, en la que el ácido nucleico que codifica cada ARN de guía es parte de un vector plasmídico o un vector vírico.

8. La composición de cualquier reivindicación precedente, en la que el ARN de guía de cada sistema CRISPR (a) se sintetiza enzimáticamente o (b) se sintetiza químicamente al menos de forma parcial.

9. Un kit que comprende la composición de una cualquiera de las reivindicaciones 1 a 8.

10. Un procedimiento *in vitro* para aumentar la eficacia y/o la especificidad de la modificación dirigida del genoma en una célula eucariota, comprendiendo el procedimiento introducir en la célula eucariota una composición que comprende:

(a) un sistema de nucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) guiado por ARN o un ácido nucleico que codifica dicho sistema de nucleasa CRISPR, en el que el sistema de nucleasa CRISPR comprende (i) una proteína de modificación de ADN programable que es una proteína CRISPR, y (ii) un ARN de guía; y

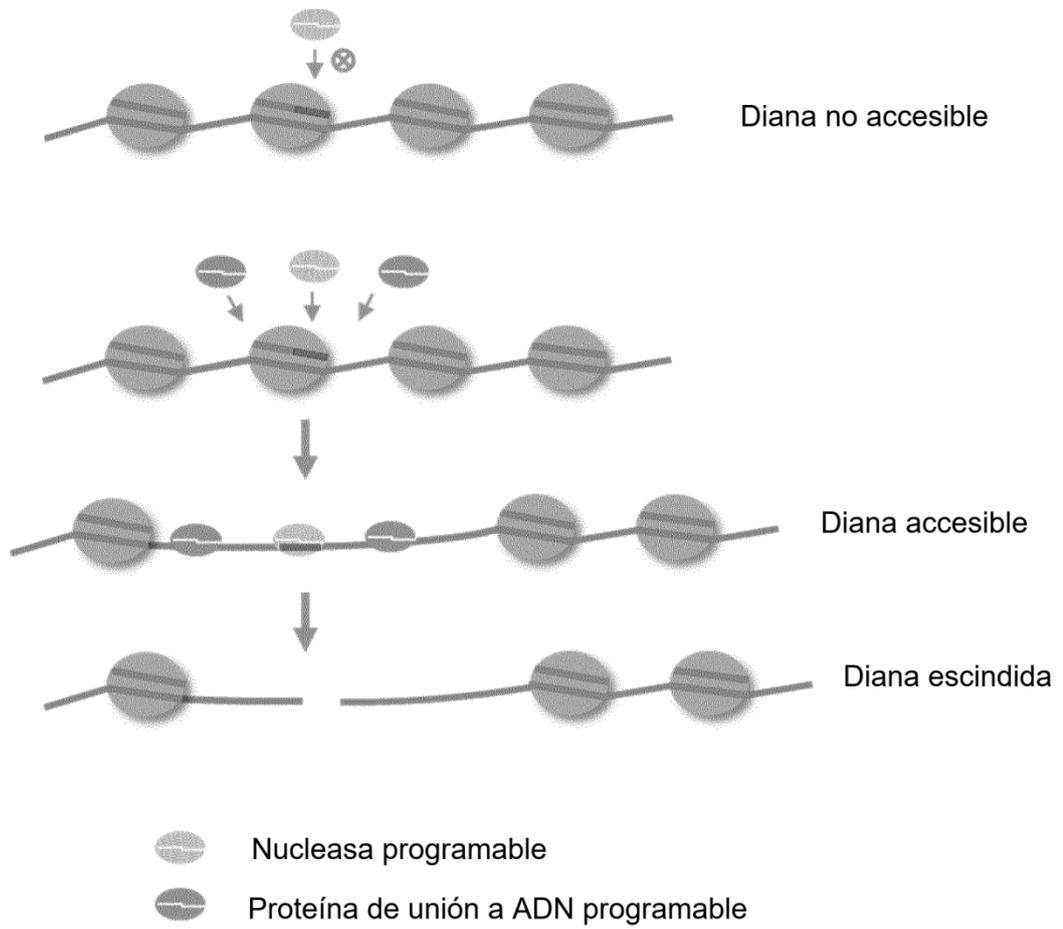
(b) al menos un sistema CRISPR catalíticamente inactivo o un ácido nucleico que codifica dicho al menos un sistema CRISPR catalíticamente inactivo, en el que cada sistema CRISPR catalíticamente inactivo comprende (i) una proteína de unión a ADN programable que es una proteína CRISPR catalíticamente inactiva y (ii) un ARN de guía; y en el que

(x) la proteína de modificación de ADN programable es una proteína CRISPR de tipo II y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II, o

(y) la proteína de modificación de ADN programable es una proteína CRISPR de tipo V y la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II;

en la que la proteína de modificación de ADN programable se dirige a una secuencia cromosómica diana y cada una de las al menos una proteína de unión a ADN programable se dirige a un sitio próximo a la secuencia cromosómica diana, y la unión de la al menos una proteína de unión a ADN programable al sitio próximo a la secuencia cromosómica diana aumenta la accesibilidad de la proteína de modificación de ADN programable a la secuencia cromosómica diana, aumentando de este modo la eficacia y/o la especificidad de la modificación dirigida del genoma.

11. El procedimiento *in vitro* de la reivindicación 10, en el que el sitio próximo a la secuencia cromosómica diana está ubicado dentro de aproximadamente 250, aproximadamente 100, aproximadamente 75, aproximadamente 50 o aproximadamente 25 pares de bases a cada lado de la secuencia cromosómica diana.
- 5 12. El procedimiento *in vitro* de la reivindicación 10 o la reivindicación 11, en el que la célula eucariota es una célula de mamífero, en el que la célula de mamífero es una célula humana o una célula no humana.
13. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 12, en el que (i) la proteína CRISPR de tipo II se selecciona de CRISPR-Cas9 de *Francisella novicida* (FnCas9), CRISPR-Cas9 de *Campylobacter jejuni* (CjCas9) y CRISPR-Cas9 de *Streptococcus pyogenes* (SpCas9); y/o (ii) la proteína CRISPR de tipo V es CRISPR-Cpf1 de *Francisella novicida* (FnCpf1).
- 10 14. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 13, en el que la al menos una proteína de unión a ADN programable es una proteína CRISPR de tipo II Cas9, en el que la proteína Cas9 tiene una o más mutaciones en cada uno del dominio de tipo RuvC y el dominio de tipo HNH.
- 15 15. El procedimiento *in vitro* de acuerdo con la reivindicación 14, en el que (i) la una o más mutaciones en el dominio de tipo RuvC son D10A, D8A, E762A y/o D986A; y/o (ii) la una o más mutaciones en el dominio de tipo HNH son H840A, H559A, N854A, N856A y/o N863A.
16. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 15, en el que el ácido nucleico que codifica cada proteína CRISPR es ARNm o ADN.
17. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 16, en el que el ácido nucleico que codifica cada proteína CRISPR es parte de un vector plasmídico o un vector vírico.
- 20 18. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 17, en el que el ácido nucleico que codifica cada ARN de guía es parte de un vector plasmídico o un vector vírico.
19. El procedimiento *in vitro* de acuerdo con cualquiera de las reivindicaciones 10 a 18, en el que el ARN de guía de cada sistema CRISPR (a) se sintetiza enzimáticamente o (b) se sintetiza químicamente al menos de forma parcial.
- 25 20. Una composición de acuerdo con cualquiera de las reivindicaciones 1-8, o el kit de acuerdo con la reivindicación 9, para su uso en terapia.
21. Una composición de acuerdo con cualquiera de las reivindicaciones 1-8, o el kit de acuerdo con la reivindicación 9, para su uso en el tratamiento de la anemia de células falciformes, talasemia, inmunodeficiencia combinada grave (IDCG), corea de Huntington o retinosis pigmentaria.



**FIG. 1**



Locus POR      SEQ ID NO: 17      diana de CjdCas9 1      Sitio de unión de SpdCas9 2

5' -CCGCTTCCCGCCTCACCCCTTGGTCTCCCTTCCAGCATTCGCCAGTACGAGCTTGTGGTCCACACCCGACATAGATGGGGCCCAAGGTGTACATGGGGGAGATGG-3'

|||||

3' -GGCGAAGGGCCGGAGTGGGAA CCGAGGGGAAAGGT CGTAA GCGGT CATGCTCGAACACCCAGGTGTGGCTGTATCTACGCGCGGTTCCACATGTACCCCTCTACC-5'

-----

Sitio de unión de SpdCas9 1      SEQ ID NO: 18

**Locus AAVS1**

SEQ ID NO: 25    5' -CAGGGATCCTGTGTCCCCGAGCTGGACCCACTTATAATCCAGGGC-3'

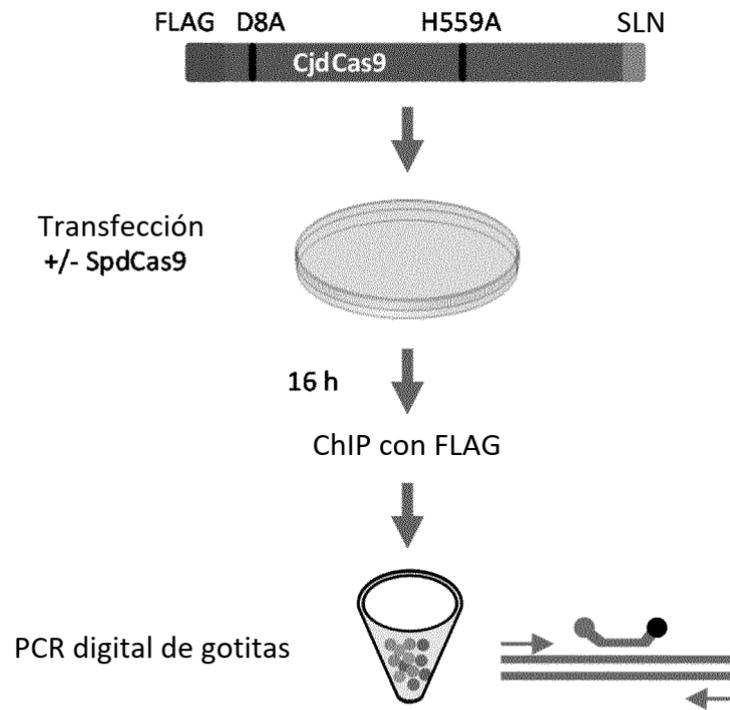
|||||

SEQ ID NO: 26    3' -GTCCCTAGGACACAGGGGCTCGACCCCTGGTGGATATAAGGGTCCCG-5'

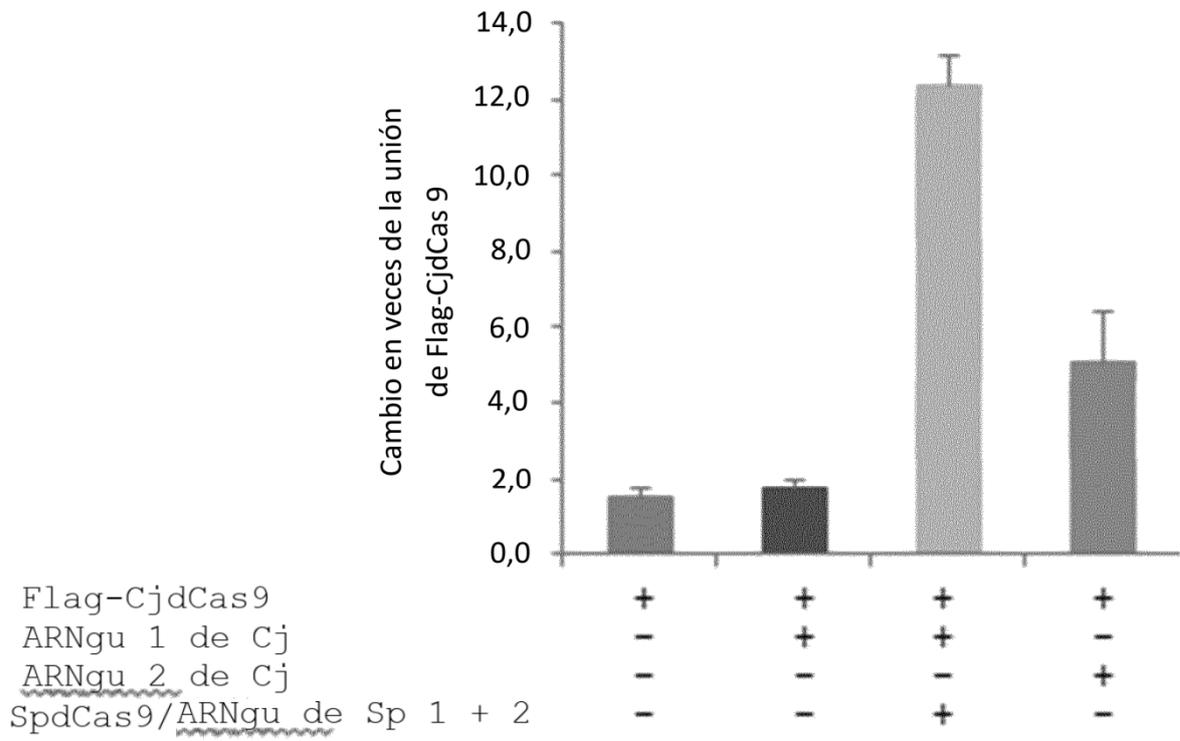
-----

diana de CjdCas9 2

**FIG. 3A**



**FIG. 3B**

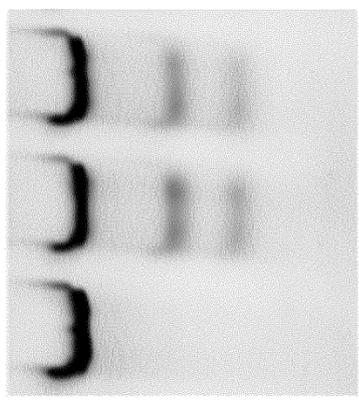


**FIG. 3C**



Locus POR  
 SEQ ID NO: 19  
 5' -CCGCTTCCCGGCTCAACCCCTGGTCTCCCTTCCAGCA**TTCCGCCAGTACGAGCTTGTGGTCCACACCCGACATAGATGCGGCCAAGGTGTACATGGGGGAGATGGCCCGG**-3'  
 3' -**GGCCGAAGGGCCCGA**GTGGGAACCCAGAGGGCAAGGTCGTAAAGCGGTCAATGCTCGAAACACCCAGGTTGGCTGTATCTACGCCCGG**TTCCACATGTACCCCTCTACCCGGCCG**-5'  
 sitio de unión de SpdCas9 1  
 sitio de unión de SpdCas9 2  
 sitio de unión de SpdCas9 3  
 diana de FnCpf1 1  
 SEQ ID NO: 20

FnCpf1	+	+	+
ARNgu de Fn	1	1	1
SpdCas9	-	+	+
Pareja de ARNgu de Sp	-	1/2	1/3



Indel (%) ND 18,7 12,6

**FIG. 5**





Locus FOR SEQ ID NO: 17 Diana de CjCas9 1 Diana de SpdCas9 2

5' -CCGCTTCCCGGCTCACCCCTTGGTCTCCCTTTCCAGCATTCGCCAGTACGAGCTTGTGTCCACACCGACATAGATGCGGCCAAAGGTGTACATGGGGGAGATGG-3'

3' -GGCGAAGGGCCGGAGTGGGAACCCAGAGGGGAAAAGTCCGTAAAGCGGTCAATGCTCCAAACACCAAGGTGTGGCTGTATCTACGCCGGTTCCACATGTACCCCTCTACC-5'

Diana de SpdCas9 1 SEQ ID NO: 18

5' -CACCCCTTGGTCTCCCTTTCCAGCATTCGCCAGTACGAGCGAATTCCTTGTGGTCCACACCGACATAGATCGGCCAAAGGTGTACATGG-3'

Donante de oligonucleótido de ADNmc Sitio de EcoRI SEQ ID NO: 27

CjCas9/ARNgu 1 de Cj + + + +

SpdCas9/ARNgu de Sp 1 + 2 - - + +

Donante de oligonucleótido de ADNmc - + + + M

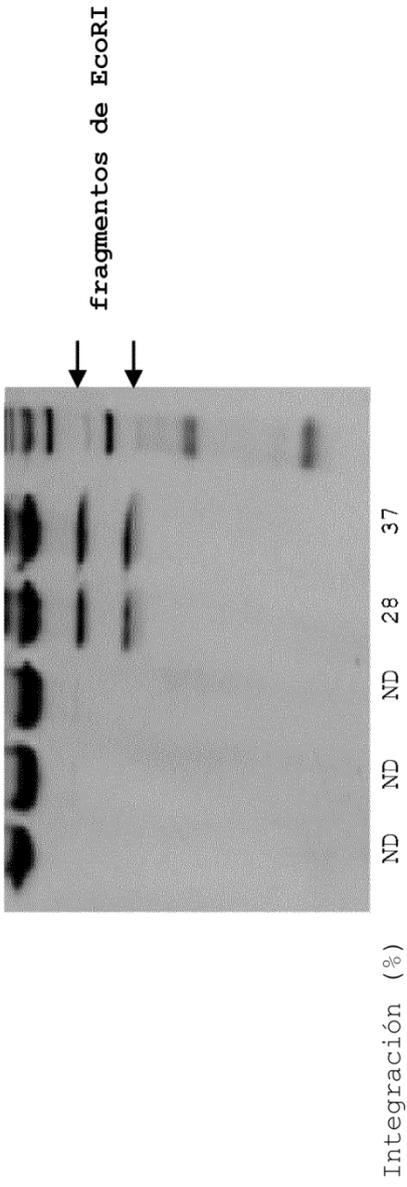


FIG. 8