

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **2 757 808**

(51) Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/63 (2006.01)

C12N 9/22 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Fecha de presentación y número de la solicitud europea: **05.12.2013 E 18156734 (8)**

(97) Fecha y número de publicación de la concesión europea: **02.10.2019 EP 3360964**

(54) Título: **Modificación y regulación del genoma basada en CRISPR**

(30) Prioridad:

06.12.2012 US 201261734256 P
30.01.2013 US 201361758624 P
05.02.2013 US 201361761046 P
15.03.2013 US 201361794422 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
30.04.2020

(73) Titular/es:

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

(72) Inventor/es:

CHEN, FUQIANG y
DAVIS, GREGORY D.

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 757 808 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

Modificación y regulación del genoma basada en CRISPR

Campo de la invención

5 La presente divulgación se refiere a la modificación del genoma dirigida. En concreto, la divulgación se refiere a endonucleasas guiadas por ARN que comprenden una proteína de tipo CRISPR/Cas y a procedimientos para usar dichas proteínas para modificar o regular secuencias cromosómicas diana.

Antecedentes de la invención

10 La modificación dirigida del genoma es una herramienta poderosa para la manipulación genética de células eucariotas, embriones y animales. Por ejemplo, pueden integrarse secuencias exógenas en ubicaciones genómicas diana y/o pueden eliminarse, inactivarse o modificarse secuencias cromosómicas endógenas específicas. Los procedimientos actuales se basan en el uso de enzimas nucleasas diseñadas por ingeniería genética, tales como, por ejemplo, nucleasas dedo de zinc (ZFN) o nucleasas efectoras tipo activador de la transcripción (TALEN). Estas nucleasas químicas contienen módulos de unión a ADN programables específicos de secuencia unidos a un dominio de escisión de ADN inespecífico. Cada nueva diana genómica, sin embargo, requiere el diseño de un nuevo ZFN o TALEN que comprenda un nuevo módulo de unión a ADN específico de secuencia. Por lo tanto, estas nucleasas diseñadas a medida tienden a ser costosas y su preparación lleva mucho tiempo. Además, las características propias de los ZFN y de los TALEN son tales que pueden mediar escisiones fuera de la diana.

20 Por lo tanto, existe la necesidad de una tecnología de modificación del genoma dirigida que no requiera el diseño de una nueva nucleasa para cada nueva ubicación genómica diana. Adicionalmente, existe la necesidad de una tecnología con mayor especificidad con pocos o ningún efecto fuera de la diana.

Sumario de la invención

Entre los diversos aspectos de la presente divulgación está la provisión de un complejo de endonucleasa guiado por ARN diseñado por ingeniería genética que comprende:

un ARN guía que comprende

25 (i) una primera región complementaria a un sitio diana en una secuencia cromosómica eucariota que puede emparejarse formando pares de bases con el sitio diana, que comprende de aproximadamente 10 nucleótidos a más de aproximadamente 25 nucleótidos,
 (ii) una segunda región que forma una estructura de tallo y bucle, y
 (iii) una tercera región que es fundamentalmente monocatenaria,

30 en el que (i), (ii) y (iii) están dispuestos en la dirección 5' a 3', y el ARN guía comprende dos moléculas separadas, en las que se forma un complejo proteína-ARN entre el ARN guía y una proteína CRISPR/Cas9 de tipo II, que comprende además una señal de localización nuclear, y el ARN guía interactúa con la proteína CRISPR/Cas9 de tipo II para guiar a la proteína al sitio diana específico. También se proporciona un ácido nucleico aislado que codifica cualquier componente de los complejos de endonucleasa guiados por ARN diseñados por ingeniería genética desvelados en el presente documento. En algunas realizaciones, el ácido nucleico puede tener codones optimizados para la traducción en células de mamífero, tales como, por ejemplo, células humanas. En otras realizaciones, la secuencia del ácido nucleico que codifica la endonucleasa guiada por ARN diseñada por ingeniería genética que forma un complejo puede unirse de forma operativa a una secuencia promotora control y, opcionalmente, puede ser parte de un vector. En otras realizaciones, un vector que comprende una secuencia que codifica la endonucleasa guiada por ARN, que puede estar unida de forma operativa a una secuencia promotora control, también puede comprender una secuencia que codifica un ARN guía, que puede estar unida de forma operativa a una secuencia promotora control.

45 Otro aspecto de la presente invención engloba el uso de dicho complejo de endonucleasa guiado por ARN diseñado por ingeniería genética y, opcionalmente, al menos un polinucleótido donador que comprende una secuencia donadora para modificar una secuencia cromosómica, en el que el uso no comprende un procedimiento para modificar la identidad genética de la línea germinal de un ser humano y, en el que dicho procedimiento no comprende un procedimiento para el tratamiento del cuerpo humano o animal mediante cirugía o terapia.

50 A continuación se detallan otros aspectos e interacciones de la divulgación.

Breve descripción de los dibujos

Figura 1 diagramas de modificación del genoma utilizando dos endonucleasas guiadas por ARN. (A) representa una ruptura bicatenaria creada por dos endonucleasas guiadas por ARN que se han convertido en nickasas. (B) representa dos roturas bicatenarias creadas por dos endonucleasas guiadas por ARN que tienen actividad

endonucleasa.

La **Figura 2** muestra la clasificación de células activadas por fluorescencia (FACS) de células K562 humanas transfectadas con el ácido nucleico de Cas9, con el ARN guía de Cas9 y con el ADN donador AAVS1-GFP. El eje Y representa la intensidad de autofluorescencia en un canal rojo y el eje X representa la intensidad de fluorescencia verde. (**A**) células K562 transfectadas con 10 µg del ARNm de Cas9 transcrita con un análogo de caperuza anti-inverso, 0,3 nmol del dúplex ARNcr-ARNtracr pre-anillado y 10 µg de ADN plasmídico de AAVS1-GFP; (**B**) células K562 transfectadas con 10 µg de ARNm de Cas9 transcrita con un análogo de caperuza anti-inverso, 0,3 nmol de ARN químérico y 10 µg de ADN plasmídico de AAVS1-GFP; (**C**) células K562 transfectadas con 10 µg de ARNm de Cas9 que se encapuchó por medio de una reacción de encapuchado postranscripcional, 0,3 nmol de ARN químérico y 10 µg de ADN plasmídico de AAVS1-GFP; (**D**) células K562 transfectadas con 10 µg de ADN plasmídico de Cas9, 5 µg de ADN plasmídico de ARN químérico de U6 y 10 µg de ADN plasmídico de AAVS1-GFP; (**E**) células K562 transfectadas con 10 µg de ADN plasmídico de AAVS1-GFP; (**F**) células K562 transfectadas solo con reactivos de transfección.

La **Figura 3** muestra un análisis de unión mediante PCR que documenta la integración dirigida del GFP en el locus AAVS1 en células humanas. Carril M: marcadores moleculares de ADN de 1 kb; Carril A: Células K562 transfectadas con 10 µg de ARNm de Cas9 transcrita con un análogo de caperuza anti-inverso, 0,3 nmol del dúplex ARNcr-ARNtracr pre-anillado y 10 µg de ADN plasmídico de AAVS1-GFP; Carril B: células K562 transfectadas con 10 µg de ARNm de Cas9 transcrita con un análogo de caperuza anti-inverso, 0,3 nmol de ARN químérico y 10 µg de ADN plasmídico de AAVS1-GFP; Carril C: células K562 transfectadas con 10 µg de ARNm de Cas9 que se encapuchó por medio de una reacción de encapuchado postranscripcional, 0,3 nmol de ARN químérico y 10 µg de ADN plasmídico de AAVS1-GFP; Carril D: células K562 transfectadas con 10 µg de ADN plasmídico de Cas9, 5 µg de ADN plasmídico de ARN químérico de U6 y 10 µg de ADN plasmídico de AAVS1-GFP; Carril E: células K562 transfectadas con 10 µg de ADN plasmídico de AAVS1-GFP; Carril F: células K562 transfectadas solo con reactivos de transfección.

25 Descripción detallada de la invención

En el presente documento se desvelan endonucleasas guiadas por ARN, que comprenden al menos una señal de localización nuclear, al menos un dominio nucleasa y al menos un dominio que interactúa con un ARN guía para dirigir a la endonucleasa a una secuencia de nucleótidos específica para la escisión. También se proporcionan ácidos nucleicos que codifican las endonucleasas guiadas por ARN, así como procedimientos para usar las endonucleasas guiadas por ARN para modificar secuencias cromosómicas de células eucariotas o embriones. La endonucleasa guiada por ARN interactúa con ARN guía específicos, cada uno de los cuales dirige la endonucleasa a un sitio diana específico, sitio en el que la endonucleasa guiada por ARN introduce una ruptura bicitánea que puede repararse mediante un proceso de reparación de ADN de modo que la secuencia cromosómica se modifica. Dado que la especificidad se proporciona por el ARN guía, la endonucleasa basada en ARN es universal y puede usarse con diferentes ARN guía para dirigirse a diferentes secuencias genómicas. Los procedimientos desvelados en el presente documento pueden usarse para dirigir y modificar secuencias cromosómicas específicas y/o introducir secuencias exógenas en ubicaciones diana en el genoma de células o embriones. Se excluyen los procedimientos que comprenden un procedimiento para modificar la identidad genética de la línea germinal de un ser humano. Además, el direccionamiento es específico con efectos limitados fuera de la diana.

40 (I) Endonucleasas guiadas por ARN

Un aspecto de la presente divulgación proporciona endonucleasas guiadas por ARN que comprenden al menos una señal de localización nuclear, que permite la entrada de la endonucleasa en los núcleos de células eucariotas y en embriones tales como, por ejemplo, embriones no humanos unicelulares. Las endonucleasas guiadas por ARN también comprenden al menos un dominio nucleasa y al menos un dominio que interactúa con un ARN guía. Un ARN guía dirige a una endonucleasa guiada por ARN a una secuencia específica de un ácido nucleico (o sitio diana). El ARN guía interactúa con la endonucleasa guiada por ARN, así como con el sitio diana de manera que, una vez dirigida al sitio diana, la endonucleasa guiada por ARN puede introducir una ruptura bicitánea en la secuencia de ácido nucleico del sitio diana. Dado que el ARN guía proporciona la especificidad para la escisión dirigida, la endonucleasa de la endonucleasa guiada por ARN es universal y puede usarse con diferentes ARN guía para escindir diferentes secuencias de ácido nucleico diana. En el presente documento se desvelan endonucleasas guiadas por ARN aisladas, ácidos nucleicos aislados (es decir, ARN o ADN) que codifican las endonucleasas guiadas por ARN, vectores que comprenden ácidos nucleicos que codifican las endonucleasas guiadas por ARN y complejos de proteína-ARN que comprenden la endonucleasa guiada por ARN más un ARN guía.

La endonucleasa guiada por ARN puede derivar de un sistema asociado a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR)/CRISPR (Cas). El sistema CRISPR/Cas puede ser un sistema de tipo I, de tipo II o de tipo III. Los ejemplos no limitantes de proteínas CRISPR/Cas adecuadas incluyen Cas3, Cas4, Cas5, Cas5e (o CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (o CasA), Cse2 (o CasB), Cse3 (o CasE), Cse4 (o CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4 y Cu1966.

La endonucleasa guiada por ARN utilizada en la presente invención deriva de un sistema CRISPR/Cas de tipo II. Más

específicamente, la endonucleasa guiada por ARN deriva de una proteína Cas9. La proteína Cas9 puede ser de *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocosphaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccii*, *Candidatus desulfuridis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthospira maxima*, *Arthospira platensis*, *Arthospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, o *Acaryochloris marina*.

En general, las proteínas CRISPR/Cas comprenden al menos un dominio de reconocimiento de ARN y/o un dominio de unión a ARN. Los dominios de reconocimiento de ARN y/o unión a ARN interactúan con los ARN guía. Las proteínas CRISPR/Cas también pueden comprender dominios nucleasa (es decir, dominios DNase o RNase), dominios de unión a ADN, dominios helicasa, dominios RNase, dominios de interacción proteína-proteína, dominios de dimerización, así como otros dominios.

La proteína tipo CRISPR/Cas puede ser una proteína CRISPR/Cas de tipo salvaje, una proteína CRISPR/Cas modificada o un fragmento de una proteína CRISPR/Cas de tipo salvaje o modificada. La proteína tipo CRISPR/Cas se puede modificar para aumentar la afinidad de unión y/o la especificidad ante un ácido nucleico, alterar una actividad enzimática y/o cambiar otra propiedad de la proteína. Por ejemplo, pueden modificarse, eliminarse o inactivarse dominios nucleasa (es decir, DNase, RNase) de la proteína tipo CRISPR/Cas.

En algunas realizaciones, la proteína tipo CRISPR/Cas puede derivar de una proteína Cas9 de tipo salvaje o de un fragmento de la misma. En otras realizaciones, la proteína tipo CRISPR/Cas puede derivar de una proteína Cas9 modificada. Por ejemplo, la secuencia de aminoácidos de la proteína Cas9 se puede modificar para alterar una o más propiedades de la proteína (por ejemplo, la actividad nucleasa, la afinidad, la estabilidad, etc.). Como alternativa, los dominios de la proteína Cas9 que no participan en la escisión guiada por ARN pueden eliminarse de la proteína de modo que la proteína Cas9 modificada es más pequeña que la proteína Cas9 de tipo salvaje.

En general, una proteína Cas9 comprende al menos dos dominios nucleasa (es decir, DNase). Por ejemplo, una proteína Cas9 puede comprender un dominio nucleasa tipo RuvC y un dominio nucleasa tipo HNH. Los dominios RuvC y HNH trabajan juntos para cortar cadenas sencillas para crear una ruptura bicanalicular en el ADN. (Jinek y col., *Science*, 337: 816-821). En algunas realizaciones, la proteína derivada de Cas9 puede modificarse para contener solo un dominio nucleasa funcional (ya sea un dominio nucleasa tipo RuvC o tipo HNH). Por ejemplo, la proteína derivada de Cas9 puede modificarse de modo que uno de los dominios nucleasa se elimine o mute de manera que ya no sea funcional (es decir, la actividad nucleasa esté ausente). En algunas realizaciones en las que uno de los dominios de nucleasa está inactivo, la proteína derivada de Cas9 es capaz de introducir un corte en un ácido nucleico bicanalicular (dicha proteína se denomina "nickasa"), pero no escinde el ADN bicanalicular. Por ejemplo, una conversión de aspartato en alanina (D10A) en un dominio tipo RuvC convierte a la proteína derivada de Cas9 en una nickasa. Asimismo, una conversión de histidina en alanina (H840A o H839A) en un dominio HNH convierte la proteína derivada de Cas9 en una nickasa. Cada dominio nucleasa puede modificarse usando procedimientos muy conocidos, tales como la mutagénesis dirigida al sitio, mutagénesis mediada por PCR y síntesis genética total, así como otros procedimientos conocidos en la técnica.

La endonucleasa guiada por ARN desvelada en el presente documento comprende al menos una señal de localización nuclear. En general, una NLS comprende un tramo de aminoácidos básicos. En la técnica se conocen señales de localización nuclear (véase, por ejemplo, Lange y col., *J. Biol. Chem.*, 2007, 282:5101-5105). Por ejemplo, en una realización, la NLS puede ser una secuencia monopartita, tal como PKKKRKV (SEQ ID NO: 1) o PKKKRRV (SEQ ID NO: 2). En otra realización, la NLS puede ser una secuencia bipartita. En otra realización más, la NLS puede ser KRPAATKKAGQAKKKK (SEQ ID NO: 3). La NLS puede ubicarse en el extremo N, en el C terminal, o en una ubicación interna de la endonucleasa guiada por ARN.

La endonucleasa guiada por ARN puede comprender además al menos un dominio de penetración celular. El dominio de penetración celular puede ser una secuencia peptídica de penetración celular derivada de la proteína TAT del VIH-1. A modo de ejemplo, la secuencia de penetración celular TAT puede ser GRKKRRQRRPPQQPKKKRKV (SEQ ID NO: 4). Como alternativa, el dominio de penetración celular puede ser TLM (PLSSIFSRIGDPPKKRKV; SEQ ID NO: 5), una secuencia peptídica de penetración celular derivada del virus de la hepatitis B humana. En otra alternativa, el dominio de penetración celular puede ser MPG (GALFLGWLGAAAGSTMAGAPKKRKV; SEQ ID NO: 6 o GALFLGFLGAAAGSTMAGAWSQPKKKRKV; SEQ ID NO: 7). En una alternativa adicional, el dominio de penetración celular puede ser Pep-1 (KETWWETWWTEWSQPKKKRKV; SEQ ID NO: 8), VP22, un péptido de penetración celular del virus Herpes simplex, o una secuencia peptídica de poliarginina. El dominio de penetración celular puede localizarse en el extremo N, en el C terminal o en una ubicación interna de la proteína.

La endonucleasa guiada por ARN también puede comprender al menos un dominio marcador. Los ejemplos no limitantes de dominios marcadores incluyen proteínas fluorescentes, etiquetas de purificación y etiquetas de epítopo. En un ejemplo, el dominio marcador puede ser una proteína fluorescente. Los ejemplos no limitantes de proteínas fluorescentes adecuadas incluyen proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, EGFP, Esmeralda, Verde Azami, Verde Azami monomérico, CopGFP, AceGFP, ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, EYFP, Citrino, Venus, YPet, PhiYFP, ZsYellow1,), proteínas fluorescentes azules (por ejemplo, EBFP, EBFP2, Azurita, mKalama1, GFPuv, Sapphire, T-sapphire,), proteínas fluorescentes cian (por ejemplo ECFP, Cerúleo, CyPet, AmCyan1, Cian de Midoriishi), proteínas fluorescentes rojas (mKate, mKate2, mPlum, DsRed monomérica, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomérica, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred) y proteínas fluorescentes naranjas (mOrange, mKO, Kusabira naranja, Kusabira naranja monomérico, mTangerine, tdTomato) o cualquier otra proteína fluorescente adecuada. En otros ejemplos, el dominio marcador puede ser una etiqueta de purificación y/o una etiqueta de epitopo. Las etiquetas ejemplares incluyen, pero sin limitarse a, glutatión-S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tiorredoxina (TRX), poli(NANP), etiqueta de purificación de afinidad en tandem (TAP), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, proteína portadora de biotina carboxilo (BCCP) y calmodulina.

Como se ha desvelado anteriormente, la endonucleasa guiada por ARN es parte de un complejo proteína-ARN que comprende un ARN guía. El ARN guía interactúa con la endonucleasa guiada por ARN para dirigir la endonucleasa a un sitio diana específico, en el que el extremo 5' del ARN guía se empareja formando pares de bases con una secuencia específica protoespaciadora.

(II) Ácidos nucleicos que codifican endonucleasas guiadas por ARN

Otro aspecto de la presente divulgación proporciona ácidos nucleicos que codifican cualquiera de las endonucleasas guiadas por ARN descritas anteriormente en la sección (I). El ácido nucleico puede ser ARN o ADN. En un ejemplo, el ácido nucleico que codifica la endonucleasa guiada por ARN es ARNm. El ARNm puede estar encapuchado en 5' y/o poliadénilado en 3'. En otro ejemplo, el ácido nucleico que codifica la endonucleasa guiada por ARN es ADN. El ADN puede estar presente en un vector (véase abajo).

El ácido nucleico que codifica la endonucleasa guiada por ARN puede tener codones optimizados para la traducción eficiente en proteína en la célula eucariota o animal de interés. Por ejemplo, los codones se pueden optimizar para la expresión en humanos, ratones, ratas, hámsteres, vacas, cerdos, gatos, perros, peces, anfibios, plantas, levaduras, insectos y similares (véase Codon Usage Database en www.kazusa.or.jp/codon/). Están disponibles programas para la optimización de codones como software gratuito (por ejemplo, OPTIMIZER en genomes.urv.es/OPTIMIZER; OptimumGene™ de GenScript en www.genscript.com/codon_opt.html). También están disponibles programas comerciales de optimización de codones.

El ADN que codifica la endonucleasa guiada por ARN puede unirse de forma operativa a al menos una secuencia promotora control. En algunas iteraciones, la secuencia codificante de ADN puede estar unida de forma operativa a una secuencia promotora control para la expresión en la célula eucariota o en el animal de interés. La secuencia promotora control puede ser constitutiva, regulada o específica de tejido. Las secuencias promotoras control constitutivas adecuadas incluyen, pero sin limitarse a, promotor temprano inmediato de citomegalovirus (CMV), promotor del virus del simio (SV40), promotor tardío principal de adenovirus, promotor del virus del sarcoma de Rous (RSV), promotor del virus del tumor mamario de ratón (MMTV), promotor de la fosfoglicerato quinasa (PGK), promotor alfa del factor de elongación (ED1), 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 promotoras control reguladas y adecuadas incluyen, sin limitarse a, las reguladas por choque térmico, metales, esteroides, antibióticos o alcohol. Los ejemplos no limitantes de promotores específicos de tejido incluyen el promotor de B29, promotor de CD14, promotor de CD43, promotor de CD45, promotor de CD68, promotor de desmina, promotor de elastasa-1, promotor de endoglolina, promotor de la fibronectina, promotor de Flt-1, promotor de GFAP, promotor de GPIIb, promotor de ICAM-2, promotor de INF-β, promotor de Mb, promotor de Nphs1, promotor de OG-2, promotor de SP-B, promotor de SYN1 y promotor de WASP. La secuencia promotora puede ser de tipo salvaje o puede modificarse para una expresión más eficiente o eficaz. En un ejemplo, el ADN codificante puede unirse de forma operativa a un promotor de CMV para la expresión constitutiva en células de mamífero.

La secuencia que codifica la endonucleasa guiada por ARN puede unirse de forma operativa a una secuencia promotora que es reconocida por una ARN polimerasa de fago para la síntesis *in vitro* de ARNm. En tales ejemplos, el ARN transcripto *in vitro* puede purificarse para su uso en los procedimientos detallados a continuación en la sección (III). Por ejemplo, la secuencia promotora puede ser una secuencia promotora T7, T3 o SP6 o una variación de una secuencia promotora T7, T3 o SP6. En una realización ejemplar, el ADN que codifica la proteína está unido de forma operativa a un promotor T7 para la síntesis de ARNm *in vitro* utilizando la ARN polimerasa T7.

En una alternativa, la secuencia que codifica la endonucleasa guiada por ARN puede unirse de forma operativa a una secuencia promotora para la expresión *in vitro* de la endonucleasa guiada por ARN en células bacterianas o eucariotas. En dichas realizaciones, la proteína expresada se puede purificar para su uso en los procedimientos detallados a continuación en la sección (III). Los promotores bacterianos adecuados incluyen, sin limitarse a, promotores T7,

promotores del operón *lac*, promotores de *trp*, variaciones de los mismos y combinaciones de los mismos. Un promotor bacteriano ejemplar es *lac*, que es un híbrido de promotores de *trp* y *lac*. Anteriormente se han enumerado ejemplos no limitantes de promotores eucariotas adecuados.

- 5 En aspectos adicionales, el ADN que codifica la endonucleasa guiada por ARN también se puede unir a una señal de poliadenilación (por ejemplo, la señal polyA de SV40, la señal polyA de la hormona de crecimiento bovino (BGH), etc.) y/o al menos una secuencia de terminación transcripcional. Adicionalmente, la secuencia que codifica la endonucleasa guiada por ARN también puede unirse a la secuencia que codifica al menos una señal de localización nuclear, al menos un dominio de penetración celular y/o al menos un dominio marcador, que están detallados anteriormente en la sección (I).
- 10 En diversas realizaciones, el ADN que codifica la endonucleasa guiada por ARN puede estar presente en un vector. Los vectores adecuados incluyen vectores plasmídicos, fágidos, cósmidos, cromosomas artificiales/minicromosomas, transposones y vectores virales (por ejemplo, vectores lentivirales, vectores virales adenoasociados, etc.). En un ejemplo, el ADN que codifica la endonucleasa guiada por ARN está presente en un vector plasmídico. Los ejemplos no limitantes de vectores plasmídicos adecuados incluyen pUC, pBR322, pET, 15 pBluescript y variantes de los mismos. El vector 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 marcadoras seleccionables (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, New York, 2003 o "Molecular Cloning: A Laboratory Manual" 20 Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3^a edición, 2001.
- En algunos ejemplos, el vector de expresión que comprende la secuencia que codifica la endonucleasa guiada por ARN puede comprender además la secuencia que codifica un ARN guía. La secuencia que codifica el ARN guía generalmente está unida de forma operativa a al menos una secuencia de control transcripcional para la expresión del ARN guía en la célula o embrión de interés. Por ejemplo, el ADN que codifica el ARN guía puede estar unido de forma operativa a una secuencia promotora que es reconocida por la ARN polimerasa III (Pol III). Los ejemplos de promotores Pol III adecuados incluyen, pero sin limitarse a, promotores de ARN de mamífero U6, U3, H1 y 7SL.

(III) Procedimiento para modificar una secuencia cromosómica usando una endonucleasa guiada por ARN

- Otro aspecto englobado por la presente invención, tal y como se indicó anteriormente, es el uso del complejo de endonucleasa guiado por ARN modificado por ingeniería genética como se desvela en el presente documento y, opcionalmente, al menos un polinucleótido donador que comprende una secuencia donadora, para modificar una secuencia cromosómica, en el que el uso no comprende un procedimiento para modificar la identidad genética de la línea germinal de un ser humano y, en el que el procedimiento no comprende un procedimiento para el tratamiento del cuerpo humano o animal mediante cirugía o terapia. En un ejemplo, la célula o el embrión se cultiva de manera que cada ARN guía dirige una endonucleasa guiada por ARN a un sitio diana en la secuencia cromosómica en la que la endonucleasa guiada por ARN introduce una ruptura bicanteraria en el sitio diana. En un ejemplo de dicho uso, la ruptura bicanteraria se repara mediante un proceso de reparación de ADN de modo que se modifica la secuencia cromosómica.

- En algunos usos, el complejo de endonucleasa guiado por ARN (o ácido nucleico codificante) se introduce en una célula o embrión, en el que o en la que la endonucleasa guiada por ARN introduce una ruptura bicanteraria en la secuencia cromosómica diana. En realizaciones en las que el polinucleótido donador opcional no está presente, la ruptura bicanteraria 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 delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la ruptura. Por consiguiente, la secuencia cromosómica diana puede modificarse o desactivarse. Por ejemplo, un solo cambio de nucleótidos (SNP) puede dar lugar a un producto proteico alterado, o un cambio en el marco de lectura de una secuencia codificante puede inactivar o "knock out" la secuencia de manera que no se produzca ningún producto proteico. En realizaciones en las que está presente el polinucleótido donador opcional, la secuencia donadora en el polinucleótido donador puede intercambiarse o integrarse en la secuencia cromosómica en el sitio diana durante la reparación de la ruptura bicanteraria. Por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por secuencias corriente arriba y corriente abajo que tienen una identidad de secuencia sustancial con secuencias corriente arriba y corriente abajo, respectivamente, con respecto al sitio diana en la secuencia cromosómica, la secuencia donadora puede intercambiarse o integrarse en la secuencia cromosómica en el sitio diana durante la reparación mediada por un proceso de reparación dirigido por homología. Como alternativa, en realizaciones en las que la secuencia donadora está flanqueada por extensiones compatibles (o las extensiones compatibles se generan *in situ* por la endonucleasa guiada por ARN), la secuencia donadora puede ligarse directamente con la secuencia cromosómica escindida mediante un proceso de reparación no homólogo durante la reparación de la ruptura bicanteraria. El intercambio o la integración de la secuencia donadora en la secuencia cromosómica modifica la secuencia cromosómica diana o introduce una secuencia exógena en la secuencia cromosómica de la célula o del embrión.

- 60 En otros ejemplos, el procedimiento puede comprender la introducción de dos endonucleasas guiadas por ARN (o

ácido nucleico codificante) y dos ARN guía (o ADN codificante) en una célula o embrión no humano, en el que las endonucleasas guiadas por ARN introducen dos rupturas bicatenarias en la secuencia cromosómica. Véase la **Figura 1B**. Las dos rupturas pueden estar en varios pares de bases, en decenas de pares de bases o pueden estar separadas por muchos miles de pares de bases. En realizaciones en las que el polinucleótido donador opcional no está presente, las roturas bicatenarias resultantes pueden repararse mediante un proceso de reparación no homólogo de modo que la secuencia entre los dos sitios de escisión se pierda y/o pueden producirse delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la(s) ruptura(s). En realizaciones en las que está presente el polinucleótido donador opcional, la secuencia donadora en el polinucleótido donador puede intercambiarse o integrarse en la secuencia cromosómica durante la reparación de las roturas bicatenarias mediante un proceso de reparación basado en la homología (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por secuencias corriente arriba y corriente abajo que tienen una identidad de secuencia sustancial con secuencias corriente arriba y corriente abajo, respectivamente, con respecto a los sitios diana en la secuencia cromosómica) o un proceso de reparación no homólogo (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por extensiones compatibles).

En otras realizaciones más, el uso puede comprender la introducción de una endonucleasa guiada por ARN modificada para escindir una hebra de una secuencia bicatenaria (o ácido nucleico codificante) y dos ARN guía (o ADN codificante) en una célula o embrión, en el que cada ARN guía dirige a la endonucleasa guiada por ARN a un sitio diana específico, sitio en el que la endonucleasa modificada escinde una cadena (es decir, corta) de la secuencia cromosómica bicatenaria, y en el que los dos cortes están en posiciones opuestas y lo suficientemente cerca como para constituir una ruptura bicatenaria. Véase la Figura 1A. En realizaciones en las que el polinucleótido donador opcional no está presente, la ruptura bicatenaria resultante puede repararse mediante un proceso de reparación no homólogo de tal manera que pueden producirse delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la ruptura. En realizaciones en las que está presente el polinucleótido donador opcional, la secuencia donadora en el polinucleótido donador puede intercambiarse o integrarse en la secuencia cromosómica durante la reparación de la ruptura bicatenaria mediante un proceso de reparación basado en homología (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por secuencias corriente arriba y corriente abajo que tienen una identidad de secuencia sustancial con secuencias corriente arriba y corriente abajo, respectivamente, con respecto a los sitios diana en la secuencia cromosómica) o un proceso de reparación no homólogo (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por extensiones compatibles).

(a) Endonucleasa guiada por ARN

El uso comprende introducir en un embrión celular o no humano al menos una endonucleasa guiada por ARN que comprende al menos una señal de localización nuclear o un ácido nucleico que codifica al menos una endonucleasa guiada por ARN que comprende al menos una señal de localización nuclear. Dichas endonucleasas guiadas por ARN y ácidos nucleicos que codifican las endonucleasas guiadas por ARN se describen anteriormente en las secciones (I) y (II), respectivamente. Sin embargo, los usos reivindicados excluyen aquellos que comprenden un procedimiento para modificar la identidad genética de la línea germinal de un ser humano.

En algunas realizaciones, la endonucleasa guiada por ARN puede introducirse en la célula o en el embrión como una proteína aislada. En dichas realizaciones, la endonucleasa guiada por ARN puede comprender además al menos un dominio de penetración celular, que facilita la captación celular de la proteína. En otras realizaciones, la endonucleasa guiada por ARN puede introducirse en la célula o en el embrión como una molécula de ARNm. En otras realizaciones más, la endonucleasa guiada por ARN puede introducirse en la célula o en el embrión como una molécula de ADN. En general, la secuencia de ADN que codifica la proteína está unida de forma operativa a una secuencia promotora que funcionará en la célula o embrión de interés. La secuencia de ADN puede ser lineal o la secuencia de ADN puede ser parte de un vector. En otras realizaciones más, la proteína se puede introducir en la célula o en el embrión como un complejo ARN-proteína que comprende la proteína y el ARN guía.

En realizaciones alternativas, el ADN que codifica la endonucleasa guiada por ARN puede comprender además una secuencia que codifica un ARN guía. En general, cada una de las secuencias que codifican la endonucleasa guiada por ARN y el ARN guía está unida de forma operativa a la secuencia promotora control apropiada que permite la expresión de la endonucleasa guiada por ARN y del ARN guía, respectivamente, en la célula o en el embrión. La secuencia de ADN que codifica la endonucleasa guiada por ARN y el ARN guía puede comprender además secuencias de control de la expresión, reguladoras y/o de procesamiento adicionales. La secuencia de ADN que codifica la endonucleasa guiada por ARN y el ARN guía puede ser lineal o puede ser parte de un vector.

(b) ARN guía

El procedimiento también comprende introducir en una célula o en un embrión al menos un ARN guía o un ADN que codifica al menos un ARN guía. Un ARN guía interactúa con la endonucleasa guiada por ARN para dirigir la endonucleasa a un sitio diana específico, sitio en el que el extremo 5' del ARN guía se empareja formando pares de bases con una secuencia protoespaciadora específica en la secuencia cromosómica.

- Cada ARN guía comprende tres regiones: una primera región en el extremo 5' que es complementaria al sitio diana en la secuencia cromosómica, una segunda región interna que forma una estructura de bucle y tallo y una tercera región 3' que permanece esencialmente monocatenaria. La primera región de cada ARN guía es diferente, de modo que cada ARN guía guía a una proteína de fusión a un sitio diana específico. La segunda y tercera regiones de cada ARN guía pueden ser las mismas en todos los ARN guía.
- La primera región del ARN guía es complementaria a la secuencia (es decir, la secuencia protoespaciadora) en el sitio diana en la secuencia cromosómica, de modo que la primera región del ARN guía puede emparejarse formando pares de bases con el sitio diana. La primera región del ARN guía comprende de aproximadamente 10 nucleótidos a más de aproximadamente 25 nucleótidos. Por ejemplo, la región de emparejamiento de bases entre la primera región del ARN guía y el sitio diana en la secuencia cromosómica puede ser de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25 o más de 25 nucleótidos de longitud. En una realización ejemplar, la primera región del ARN guía tiene aproximadamente 19, 20 o 21 nucleótidos de longitud.
- El ARN guía también comprende una segunda región que forma una estructura secundaria. En algunas realizaciones, la estructura secundaria comprende un tallo (o horquilla) y un bucle. La longitud del bucle y el tallo pueden 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 una 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. En una realización ejemplar, el bucle tiene aproximadamente 4 nucleótidos de longitud y el tallo comprende aproximadamente 12 pares de bases.
- El ARN 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 cromosómica en la célula de interés y no tiene complementariedad con el resto del ARN guía. La longitud de la tercera región puede variar. En general, la tercera región tiene 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 la segunda y tercera regiones (también llamada región universal o de armazón) del ARN guía puede variar de aproximadamente 30 a aproximadamente 120 nucleótidos de longitud. En un aspecto, la longitud combinada de las regiones segunda y tercera del ARN guía varía de aproximadamente 70 a aproximadamente 100 nucleótidos de longitud.
- El ARN guía comprende dos moléculas separadas. La primera molécula de ARN puede comprender la primera región del ARN guía y la mitad del "tallos" de la segunda región del ARN guía. La segunda molécula de ARN puede comprender la otra mitad del "tallos" de la segunda región del ARN guía y la tercera región del ARN guía. Por lo tanto, en esta realización, la primera y segunda moléculas de ARN contienen cada una una secuencia de nucleótidos que son complementarias entre sí. Por ejemplo, en una realización, la primera y segunda moléculas de ARN comprenden cada una una secuencia (de aproximadamente 6 a aproximadamente 20 nucleótidos) que se empareja formando pares de bases con la otra secuencia para formar un ARN guía funcional.
- En algunas realizaciones, el ARN guía se puede introducir en la célula o en el embrión como una molécula de ARN. La molécula de ARN se puede transcribir *in vitro*. Como alternativa, la molécula de ARN puede sintetizarse químicamente.
- En otras realizaciones, el ARN guía puede introducirse en la célula o en el embrión como una molécula de ADN. En tales casos, el ADN que codifica el ARN guía puede estar unido de forma operativa a la secuencia promotora control para la expresión del ARN guía en la célula o embrión de interés. Por ejemplo, la secuencia codificante del ARN puede estar unida de forma operativa a una secuencia promotora que es reconocida por la ARN polimerasa III (Pol III). Los ejemplos de promotores Pol III adecuados incluyen, pero sin limitarse a, promotores U6 o H1 de mamíferos. En realizaciones ejemplares, la secuencia codificante del ARN está unida a un promotor U6 de ratón o humano. En otras realizaciones ejemplares, la secuencia codificante del ARN está unida a un promotor H1 de ratón o humano.
- La molécula de ADN que codifica el ARN guía puede ser lineal o circular. En algunas realizaciones, la secuencia de ADN que codifica el ARN guía puede ser parte de un vector. Los vectores adecuados incluyen vectores plasmídicos, fágimidos, cósmidos, cromosomas artificiales/minicromosomas, transposones y vectores virales. En una realización ejemplar, el ADN que codifica la endonucleasa guiada por ARN está presente en un vector plasmídico. Los ejemplos no limitantes de vectores plasmídicos adecuados incluyen pUC, pBR322, pET, pBluescript y variantes de los mismos. El vector 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 marcadoras seleccionables (por ejemplo, genes de resistencia a antibióticos), orígenes de replicación y similares.
- En realizaciones en las que tanto la endonucleasa guiada por ARN como el ARN guía se introducen en la célula como moléculas de ADN, cada una puede ser parte de una molécula separada (por ejemplo, un vector que contiene la secuencia de codificación de proteínas y un segundo vector que contiene la secuencia de codificación del ARN guía) o ambas pueden ser parte de la misma molécula (por ejemplo, un vector que contiene la secuencia de codificación (y

reguladora) tanto para la proteína como para el ARN guía).

(c) Sitio diana

Una endonucleasa guiada por ARN junto con un ARN guía se dirige a un sitio diana en la secuencia cromosómica, en la que la endonucleasa guiada por ARN introduce una ruptura bicatenaria en la secuencia cromosómica. El sitio diana no tiene limitación de secuencia, excepto que la secuencia va seguida inmediatamente (corriente abajo) por una secuencia de consenso. Esta secuencia consenso también se conoce como motivo adyacente al protoespaciador (PAM). Los ejemplos de PAM incluyen, pero sin limitarse a, NGG, NGGNG y NNAGAAW (en los que N se define como cualquier nucleótido y W se define como A o T). Como se detalla anteriormente en la sección (III) (b), la primera región (en el extremo 5') del ARN guía es complementaria al protoespaciador de la secuencia diana. Típicamente, la primera región del ARN guía tiene una longitud de aproximadamente 19 a 21 nucleótidos. Por lo tanto, en determinados aspectos, la secuencia del sitio diana en la secuencia cromosómica es 5'-N₁₉₋₂₁-NGG-3'. El PAM está en cursiva.

El sitio diana puede estar en la región codificante de un gen, en un intrón de un gen, en una región de control de un gen, en una región no codificante entre genes, etc. El gen puede ser un gen codificante de proteínas o un gen codificante de ARN. El gen puede ser cualquier gen de interés.

15 (d) Polinucleótido donador opcional

En algunas realizaciones, el uso comprende además introducir al menos un polinucleótido donador en la célula o el embrión. Un polinucleótido donador comprende al menos una secuencia donadora. En algunos aspectos, una secuencia donadora del polinucleótido donador corresponde a una secuencia cromosómica endógena o nativa. Por ejemplo, la secuencia donadora puede ser esencialmente idéntica a una porción de la secuencia cromosómica en o cerca del sitio diana, pero que comprende al menos un cambio de nucleótido. Por lo tanto, la secuencia donadora puede comprender una versión modificada de la secuencia de tipo salvaje en el sitio diana de manera que, tras la integración o el intercambio con la secuencia nativa, la secuencia en la ubicación cromosómica diana 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 las mismas. Como consecuencia de la integración de la secuencia modificada, la célula o el embrión/animal pueden producir un producto génico modificado a partir de la secuencia cromosómica diana.

En otros aspectos, la secuencia donadora del polinucleótido donador corresponde a una secuencia exógena. Tal como se usa en el presente documento, una secuencia "exógena" se refiere a una secuencia que no es nativa de la célula o del embrión, o una secuencia cuya ubicación nativa en el genoma de la célula o el embrión está en una ubicación diferente. Por ejemplo, la secuencia exógena puede comprender una secuencia codificante de proteínas, que se puede unir de forma operativa a una secuencia promotora control exógena de manera que, tras la integración en el genoma, la célula o el embrión/animal es capaz de expresar la proteína codificada por la secuencia integrada. Como alternativa, la secuencia exógena puede integrarse en la secuencia cromosómica de manera que su expresión esté regulada por una secuencia promotora control endógena. En otras iteraciones, la secuencia exógena puede ser una secuencia de control transcripcional, otra secuencia de control de la expresión, una secuencia de codificación de ARN, y así sucesivamente. La integración de una secuencia exógena en una secuencia cromosómica se denomina "knock in".

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

40 Polinucleótido donador que comprende secuencias corriente arriba y corriente abajo. En algunas realizaciones, la secuencia donadora en el polinucleótido donador está flanqueada por una secuencia corriente arriba y una secuencia corriente abajo, que tienen identidad de secuencia sustancial con respecto a secuencias ubicadas corriente arriba y corriente abajo, respectivamente, con respecto al sitio diana en la secuencia cromosómica. Debido a estas similitudes de secuencia, las secuencias corriente arriba y corriente abajo del polinucleótido donador permiten la recombinación homóloga entre el polinucleótido donador y la secuencia cromosómica diana de manera que la secuencia del donador puede integrarse (o intercambiarse) con la secuencia cromosómica.

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 del sitio diana. De forma similar, la secuencia corriente abajo se refiere a una secuencia de ácido nucleico que comparte una identidad de secuencia sustancial con una secuencia cromosómica corriente abajo del sitio diana. Tal como se usa en el presente documento, la frase "identidad de secuencia sustancial" se refiere a secuencias que tienen al menos aproximadamente un 75 % de identidad de secuencia. Por lo tanto, las secuencias corriente arriba y corriente abajo en el polinucleótido donante pueden tener aproximadamente un 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 secuencias corriente arriba o corriente abajo del sitio diana. En una realización ejemplar, las secuencias corriente arriba y corriente abajo en el polinucleótido donador pueden tener una identidad de secuencia de aproximadamente un 95 % o 100 % con secuencias cromosómicas corriente arriba o corriente abajo del sitio diana. En una realización, la secuencia corriente arriba comparte una identidad de secuencia sustancial con una secuencia

cromosómica ubicada inmediatamente corriente arriba del sitio diana (es decir, adyacente al sitio diana). En otras realizaciones, la secuencia corriente arriba comparte una identidad de secuencia sustancial con una secuencia cromosómica que se encuentra dentro de aproximadamente cien (100) nucleótidos corriente arriba del sitio diana. Por lo tanto, por ejemplo, la secuencia corriente arriba puede compartir una identidad de secuencia sustancial con una secuencia cromosómica que se encuentra 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 del sitio diana. En una realización, la secuencia corriente abajo comparte una identidad de secuencia sustancial con una secuencia cromosómica ubicada inmediatamente corriente abajo del sitio diana (es decir, adyacente al sitio objetivo). En otras realizaciones, la secuencia corriente abajo comparte una identidad de secuencia sustancial con una secuencia cromosómica que se encuentra dentro de aproximadamente cien (100) nucleótidos corriente abajo del sitio diana. Por lo tanto, por ejemplo, la secuencia corriente abajo puede compartir una identidad de secuencia sustancial con una secuencia cromosómica que se encuentra 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 del sitio diana.

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 ejemplares, las secuencias corriente arriba y corriente abajo pueden variar en longitud de aproximadamente 50 a aproximadamente 1500 nucleótidos.

Los polinucleótidos donadores que comprenden las secuencias corriente arriba y corriente abajo con similitud de secuencia con la secuencia cromosómica diana pueden ser lineales o circulares. En realizaciones en las que el polinucleótido donador es circular, este puede ser parte de un vector. Por ejemplo, el vector puede ser un vector plasmídico.

Polinucleótido donador que comprende uno o más sitios de escisión. En otras realizaciones, el polinucleótido donador puede comprender adicionalmente al menos un sitio de escisión diana que es reconocido por la endonucleasa guiada por ARN. El sitio de escisión diana agregado al polinucleótido donador puede colocarse corriente arriba o corriente abajo o tanto corriente arriba como corriente abajo de la secuencia donadora. Por ejemplo, la secuencia donadora puede estar flanqueada por sitios de escisión diana de manera que, tras la escisión por la endonucleasa guiada por ARN, la secuencia donadora está flanqueada por extensiones que son compatibles con las de la secuencia cromosómica generada tras la escisión por la endonucleasa guiada por ARN. Por consiguiente, la secuencia donadora puede ligarse con la secuencia cromosómica escindida durante la reparación de la ruptura biciatenaria mediante un proceso de reparación no homólogo. En general, los polinucleótidos donadores que comprenden el o los sitios de escisión diana serán circulares (por ejemplo, pueden ser parte de un vector plasmídico).

Polinucleótido donador que comprende una secuencia donadora corta con extensiones opcionales. En otras realizaciones alternativas, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta con extensiones cortas opcionales que son compatibles con las extensiones generadas por la endonucleasa guiada por ARN. En dichas realizaciones, la secuencia donadora se puede ligar directamente con la secuencia cromosómica escindida durante la reparación de la ruptura biciatenaria. En algunas ocasiones, la secuencia donadora puede ser inferior a aproximadamente 1000, inferior a aproximadamente 500, inferior a aproximadamente 250 o inferior a aproximadamente 100 nucleótidos. En ciertos casos, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta con extremos romos. En otras iteraciones, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta con extensiones 5' y/o 3'. Las extensiones pueden comprender 1, 2, 3, 4 o 5 nucleótidos.

Típicamente, el polinucleótido donador será ADN. El ADN puede ser monocatenario o biciatenario y/o lineal o circular. El polinucleótido donador puede ser un plásmido de ADN, un cromosoma artificial bacteriano (BAC), un cromosoma artificial de levadura (YAC), un vector viral, una pieza lineal de ADN, un fragmento de PCR, un ácido nucleico desnudo, o un ácido nucleico que forma un complejo con un vehículo de suministro tal como un liposoma o poloxámero. En ciertas realizaciones, el polinucleótido donador que comprende la secuencia donadora puede ser parte de un vector plasmídico. En cualquiera de estas situaciones, el polinucleótido donador que comprende la secuencia donadora puede comprender además al menos una secuencia adicional.

(e) Introducción en la célula o embrión

La o las endonucleasas dirigidas por ARN (o ácido nucleico codificante), el o los ARN guía (o ADN codificante), y el o los polinucleótidos donadores opcionales pueden introducirse en una célula o embrión por una variedad de medios. En algunas realizaciones, se transfetta la célula o el embrión. Los procedimientos de transfección adecuados incluyen transfección mediada por fosfato de calcio, nucleofección (o electroporación), transfección de polímeros catiónicos (por ejemplo, DEAE-dextrano o polietilenimina), transducción viral, transfección con virosomas, transfección con viriones, transfección con liposomas, transfección con liposomas catiónicos, transfección con inmunoliposomas, transfección con lípidos no liposomales, transfección con dendrímeros, transfección por choque térmico,

magnetofección, lipofección, biobalística génica, impalefección, sonicación, 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, New York, 2003 o "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3^a edición, 2001).

5 En otras realizaciones, las moléculas se introducen en la célula o en el embrión por microinyección. Típicamente, el embrión es un embrión fertilizado en la fase unicelular de la especie de interés. Por ejemplo, las moléculas pueden inyectarse en los pronúcleos de embriones unicelulares.

10 La o las endonucleasas dirigidas por ARN (o ácido nucleico codificante), el o los ARN guía (o los ADN que codifican el ARN guía), y el o los polinucleótidos donadores opcionales pueden introducirse en la célula o en el embrión simultáneamente o secuencialmente. La relación de la(s) endonucleasa(s) dirigida(s) al ARN (o ácido nucleico codificante) al(las) ARN guía (o ADN codificante) generalmente será aproximadamente estequiométrica de modo que puedan formar un complejo ARN-proteína. En una realización, el ADN que codifica una endonucleasa dirigida por ARN y el ADN que codifica un ARN guía se suministran juntos dentro del vector plasmídico.

(f) Cultivo de la célula o el embrión

15 El uso comprende además mantener la célula o el embrión en condiciones apropiadas de modo que el o los ARN guía dirigen la o las endonucleasas guiadas por el ARN al sitio o sitios diana en la secuencia cromosómica, y la o las endonucleasas guiadas por el ARN introducen al menos una ruptura bicondicional en la secuencia cromosómica. Una ruptura bicondicional puede repararse mediante un proceso de reparación del ADN de modo que la secuencia cromosómica se modifique mediante la delección de al menos un nucleótido, una inserción de al menos un nucleótido, 20 una sustitución de al menos un nucleótido, o una combinación de las mismas.

25 En realizaciones en las que no se introduce un polinucleótido donador en la célula o en el embrión, la ruptura bicondicional se puede reparar mediante un proceso de reparación de unión de extremos no homólogos (NHEJ). Debido a que el NHEJ es propenso a errores, pueden producirse delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la ruptura. Por consiguiente, la secuencia en la secuencia cromosómica se puede modificar de modo que el marco de lectura de una región codificante se pueda desplazar y que la secuencia cromosómica se inactive o se someta a "knock out". Una secuencia cromosómica codificante de la proteína inactivada no da lugar a la proteína codificada por la secuencia cromosómica de tipo salvaje.

30 En realizaciones en las que se introduce un polinucleótido donador que comprende secuencias corriente arriba y corriente abajo en la célula o en el embrión, la ruptura bicondicional puede repararse mediante un proceso de reparación dirigida por homología (HDR) de modo que la secuencia donadora se integre en la secuencia cromosómica. Por consiguiente, una secuencia exógena puede integrarse en el genoma de la célula o en el embrión, o la secuencia cromosómica diana puede modificarse mediante el intercambio de una secuencia modificada por la secuencia cromosómica de tipo salvaje.

35 En realizaciones en las que se introduce un polinucleótido donador que comprende el sitio de escisión diana en la célula o en el embrión, la endonucleasa guiada por ARN puede escindir tanto la secuencia cromosómica diana como el polinucleótido donador. El polinucleótido donador linealizado puede integrarse en la secuencia cromosómica en el sitio de la ruptura bicondicional mediante ligación entre el polinucleótido donador y la secuencia cromosómica escindida a través de un proceso NHEJ.

40 En realizaciones en las que un polinucleótido donador lineal que comprende una secuencia donadora corta se introduce en la célula o en el embrión, la secuencia donadora corta puede integrarse en la secuencia cromosómica en el sitio de la ruptura bicondicional mediante un proceso NHEJ. La integración puede llevarse a cabo mediante la ligación de extremos romos entre la secuencia corta donadora y la secuencia cromosómica en el sitio de la ruptura bicondicional. Como alternativa, la integración puede llevarse a cabo mediante ligación de extremos cohesivos (es decir, que tienen extensiones 5' o 3') entre una secuencia corta donadora que está flanqueada por extensiones que son compatibles con las generadas por la endonucleasa dirigida por ARN en la secuencia cromosómica escindida.

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

50 Un embrión se puede cultivar *in vitro* (por ejemplo, en cultivo celular). Típicamente, el embrión se cultiva a una temperatura apropiada y en medios apropiados con la proporción O₂/CO₂ necesaria para permitir la expresión de la endonucleasa de ARN y del ARN guía, si es necesario. Los ejemplos de medios no limitantes adecuados incluyen los medios M2, M16, KSOM, BMOC y HTF. Un experto en la materia apreciará que las condiciones de cultivo pueden variar y variarán según las especies del embrión. En todos los casos, se puede usar la optimización rutinaria, para

determinar las mejores condiciones de cultivo para una especie particular de embrión. En algunos casos, una línea celular puede derivar de un embrión cultivado *in vitro*(por ejemplo, una línea de células madre embrionarias).

Como alternativa, un embrión puede cultivarse *in vivo* transfiriendo el embrión al útero de un hospedador femenino. En términos generales, el hospedador femenino es de la misma especie o de una especie similar a la del embrión.

- 5 Preferentemente, la hembra hospedadora está pseudoembarazada. En la técnica se conocen procedimientos para preparar huéspedes hembra pseudoembarazadas. Adicionalmente, se conocen procedimientos para transferir un embrión a un hospedador femenino. Cultivar un embrión *in vivo* permite que el embrión se desarrolle y pueda dar lugar a un nacimiento vivo de un animal derivado del embrión. Dicho animal comprendería la secuencia cromosómica modificada en cada célula del cuerpo. Los usos que comprenden un procedimiento para modificar la identidad genética 10 de la línea germinal de un ser humano están específicamente excluidos del ámbito de la invención.

(g) Tipos de células y embriones

Una variedad de células eucariotas y embriones no humanos son adecuados para el uso desvelado en el presente documento. Por ejemplo, la célula puede ser una célula humana, 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. En general, el embrión es un embrión de mamífero no humano. En realizaciones específicas, los embriones pueden ser un embrión unicelular de mamífero no humano. Los embriones de mamíferos ejemplares, incluyendo embriones unicelulares, incluyen sin limitarse a, embriones de ratón, rata, hámster, roedor, conejo, felino, canino, ovino, porcino, bovino, equino y de primate. En otras realizaciones más, la célula puede ser una célula madre. Las células madre adecuadas incluyen, sin limitarse a, células madre embrionarias, células madre tipo ES, células madre fetales, células madre adultas, células madre pluripotentes, células madre pluripotentes inducidas, células madre multipotentes, células madre oligopotentes, células madre unipotentes y otras. En realizaciones ejemplares, la célula es una célula de mamífero.

Ejemplos no limitantes de células de mamífero adecuadas incluyen células de ovario de hámster chino (CHO), células de riñón de cría de hámster (BHK); células de mieloma de ratón NS0, células de fibroblastos embrionarios de ratón 25 3T3 (NIH3T3), células de linfoma de linfocitos B de ratón A20; células de melanoma de ratón B16; células de mioblastos de ratón C2C12; células de mieloma de ratón SP2/0; células mesenquimales embrionarias de ratón C3H-10T1/2; células de carcinoma de ratón CT26, células de próstata de ratón DuCuP; células de mama de ratón EMT6; células de hepatoma de ratón Hepa1c1c7; células de mieloma de ratón J5582; células epiteliales de ratón MTD-1A; células de miocardio del ratón MyEnd; células renales de ratón RenCa; células pancreáticas de ratón RIN-5F; células de melanoma de ratón X64; células de linfoma de ratón YAC-1; células de glioblastoma de rata 9L; células de linfoma de 30 linfocitos B de rata RBL; células de neuroblastoma de rata B35; células de hepatoma de rata (HTC); células de hígado de rata búfalo BRL 3A; células de riñón canino (MDCK); células mamarias caninas (CMT); células de osteosarcoma de rata D17; células de monocitos/macrófagos de rata DH82; células de fibroblastos transformados con SV-40 de riñón 35 de mono (COS7); células de riñón de mono CVI-76; células de riñón de mono verde africano (VERO-76); células de riñón embrionario humano (HEK293, HEK293T); células de carcinoma cervical humano (HELA); células de pulmón humano (W138); células hepáticas humanas (Hep G2); células de osteosarcoma humano U2-OS, células humanas A549, células humanas A-431 y células humanas K562. Se puede encontrar una extensa lista de líneas celulares de mamíferos en el catálogo de la American Type Culture Collection (ATCC, Manassas, VA).

(IV) Células y animales genéticamente modificados

40 La presente divulgación engloba células genéticamente modificadas, embriones no humanos y animales no humanos que comprenden al menos una secuencia cromosómica que se ha modificado usando un complejo de endonucleasa guiada por ARN diseñado por ingeniería genética descrito en el presente documento. La divulgación proporciona células que comprenden al menos una molécula de ADN o ARN que codifica una endonucleasa guiada por ARN dirigida a una secuencia cromosómica de interés, al menos un ARN guía y, opcionalmente, uno o más polinucleótidos 45 donadores. La divulgación también proporciona embriones no humanos que comprenden al menos una molécula de ADN o ARN que codifica una endonucleasa guiada por ARN dirigida a una secuencia cromosómica de interés, al menos un ARN guía y, opcionalmente, uno o más polinucleótidos donadores.

La presente divulgación proporciona animales no humanos genéticamente modificados, embriones no humanos o 50 células animales que comprenden al menos una secuencia cromosómica modificada. La secuencia cromosómica modificada puede modificarse de modo que (1) esté inactivada, (2) tenga una expresión alterada o produzca un producto proteico alterado o (3) comprenda una secuencia integrada. La secuencia cromosómica se modifica con un procedimiento guiado por ARN mediado por endonucleasa o por proteína de fusión, utilizando los procedimientos descritos en el presente documento.

55 Como se ha tratado, un aspecto de la presente divulgación proporciona un animal genéticamente modificado en el que se ha modificado al menos una secuencia cromosómica. En una realización, el animal modificado genéticamente comprende al menos una secuencia cromosómica inactivada. La secuencia cromosómica modificada puede inactivarse de modo que la secuencia no se transcriba y/o no se produzca un producto proteico funcional. Por lo tanto, un animal genéticamente modificado que comprende una secuencia cromosómica inactivada puede denominarse "knock out" o "knock out condicional". La secuencia cromosómica inactivada puede incluir una mutación de delección

(es decir, delección de uno o más nucleótidos), una mutación de inserción (es decir, inserción de uno o más nucleótidos), o una mutación sin sentido (es decir, la sustitución de un único nucleótido por otro nucleótido de manera que se introduce un codón de parada). Como consecuencia de la mutación, la secuencia cromosómica diana se inactiva y no se produce una proteína funcional. La secuencia cromosómica inactivada comprende una secuencia introducida de manera no exógena. También se incluyen en el presente documento animales genéticamente modificados en los que dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o diez o más secuencias cromosómicas están inactivadas.

En otra realización, la secuencia cromosómica modificada puede alterarse de modo que codifique una variante proteica como producto. Por ejemplo, un animal genéticamente modificado que comprende una secuencia cromosómica modificada puede comprender una o más mutaciones puntuales u otra modificación de manera que se produzca un producto proteico alterado. En una realización, la secuencia cromosómica se puede modificar de modo que al menos un nucleótido se cambie y la proteína expresada comprenda un resto de aminoácido modificado (mutación de sentido erróneo). En otra realización, la secuencia cromosómica puede modificarse para comprender más de una mutación de sentido erróneo de modo que se cambie más de un aminoácido. Adicionalmente, la secuencia cromosómica se puede modificar para que tenga una delección o inserción de tres nucleótidos de modo que la proteína expresada comprenda una delección o inserción de un solo aminoácido. La proteína alterada o variante puede tener propiedades o actividades alteradas en comparación con la proteína de tipo salvaje, tales como la especificidad de sustrato alterada, actividad enzimática alterada, tasas cinéticas alteradas, etc.

En otra realización, el animal modificado genéticamente puede comprender al menos una secuencia cromosómicamente integrada. Un animal genéticamente modificado que comprende una secuencia integrada puede denominarse "knock in" o "knock in condicional". La secuencia cromosómicamente integrada puede, por ejemplo, codificar una proteína ortóloga, una proteína endógena, o combinaciones de ambas. En una realización, una secuencia que codifica una proteína ortóloga o una proteína endógena puede integrarse en una secuencia cromosómica que codifica una proteína de modo que la secuencia cromosómica se inactive, pero la secuencia exógena se exprese. En ese caso, la secuencia que codifica la proteína ortóloga o la proteína endógena puede estar unida de forma operativa a una secuencia promotora control. Como alternativa, una secuencia que codifica una proteína ortóloga o una proteína endógena puede integrarse en una secuencia cromosómica sin afectar a la expresión de una secuencia cromosómica. Por ejemplo, una secuencia que codifica una proteína puede integrarse en un locus de "puerto seguro", como los loci Rosa26, Locus HPRT o locus AAV. La presente divulgación también engloba animales genéticamente modificados en los que dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o diez o más secuencias, incluyendo secuencias que codifican una o más proteínas, están integradas en el genoma.

La secuencia cromosómicamente integrada que codifica una proteína puede codificar la forma de tipo salvaje de una proteína de interés o puede codificar una proteína que comprende al menos una modificación tal que se produce una versión alterada de la proteína. Por ejemplo, una secuencia cromosómicamente integrada que codifica una proteína relacionada con una enfermedad o trastorno puede comprender al menos una modificación tal que la versión alterada de la proteína producida cause o potencie el trastorno asociado. Como alternativa, la secuencia cromosómicamente integrada que codifica una proteína relacionada con una enfermedad o trastorno puede comprender al menos una modificación de tal manera que la versión alterada de la proteína proteja contra el desarrollo del trastorno asociado.

En un ejemplo adicional, el animal modificado genéticamente puede ser un animal "humanizado" que comprende al menos una secuencia cromosómicamente integrada que codifica una proteína humana funcional. La proteína humana funcional puede no tener un ortólogo correspondiente en el animal genéticamente modificado. Como alternativa, el animal de tipo salvaje del que deriva el animal genéticamente modificado puede comprender un ortólogo correspondiente a la proteína humana funcional. En este caso, la secuencia ortóloga en el animal "humanizado" se inactiva de tal manera que no se produce proteína funcional y el animal "humanizado" comprende al menos una secuencia cromosómicamente integrada que codifica la proteína humana.

En otro ejemplo más, el animal modificado genéticamente puede comprender al menos una secuencia cromosómica modificada que codifica una proteína de manera que se altera el patrón de expresión de la proteína. Por ejemplo, regiones reguladoras que controlan la expresión de la proteína, como un promotor o un sitio de unión de un factor de transcripción, pueden alterarse de modo que la proteína se sobreproduce, o se altera la expresión temporal o específica de tejido de la proteína, o una combinación de las mismas. Como alternativa, el patrón de expresión de la proteína se puede alterar utilizando un sistema de knockout condicional. Un ejemplo no limitante de un sistema de knockout condicional incluye un sistema de recombinación Cre-lox. Un sistema de recombinación Cre-lox comprende una enzima recombinasa Cre, una recombinasa de ADN específica de sitio que puede catalizar la recombinación de una secuencia de ácido nucleico entre sitios específicos (sitios lox) en una molécula de ácido nucleico. Los procedimientos para usar este sistema para producir una expresión temporal y específica de tejido son conocidos en la técnica. En general, se genera un animal genéticamente modificado con sitios lox que flanquean a una secuencia cromosómica. El animal genéticamente modificado que comprende la secuencia cromosómica flanqueada por lox puede cruzarse después con otro animal genéticamente modificado que expresa la recombinasa Cre. Luego se producen animales de la progenie que comprenden la secuencia cromosómica flanqueada por lox y la recombinasa Cre, y la secuencia cromosómica flanqueada por lox se recombinan, conduciendo a la delección o inversión de la secuencia cromosómica que codifica la proteína. La expresión de la recombinasa Cre puede regularse temporal y condicionalmente para efectuar una recombinación regulada temporal y condicionalmente de la secuencia cromosómica.

En cualquiera de estas realizaciones, el animal modificado genéticamente desvelado en el presente documento puede ser heterocigoto para la secuencia cromosómica modificada. Como alternativa, el animal modificado genéticamente puede ser homocigoto para la secuencia cromosómica modificada.

- 5 Los animales genéticamente modificados desvelados en el presente documento pueden cruzarse para crear animales que comprenden más de una secuencia cromosómica modificada o para crear animales que sean homocigotos para una o más secuencias cromosómicas modificadas. Por ejemplo, dos animales que comprenden la misma secuencia cromosómica modificada pueden cruzarse para crear un animal homocigoto para la secuencia cromosómica modificada. Como alternativa, los animales con diferentes secuencias cromosómicas modificadas pueden cruzarse para crear un animal que comprenda ambas secuencias cromosómicas modificadas.
- 10 Por ejemplo, un primer animal que comprende un gen de secuencia cromosómica inactivada "x" puede cruzarse con un segundo animal que comprende una secuencia cromosómicamente integrada que codifica una proteína "X" de gen humano para dar lugar a una descendencia con el gen "X" "humanizado" que comprende tanto la secuencia cromosómica inactivada del gen "x" como la secuencia "X" del gen humano integrado cromosómicamente. Asimismo, un animal con el gen "X" humanizado puede cruzarse con un animal con el gen "Y" humanizado para crear descendencia con el gen X/gen Y humanizados. Los expertos en la materia apreciarán que son posibles muchas combinaciones.
- 15

- 20 En otras realizaciones, un animal que comprende una secuencia cromosómica modificada puede cruzarse para combinar la secuencia cromosómica modificada con otros antecedentes genéticos. A modo de ejemplo no limitativo, otros antecedentes genéticos pueden incluir antecedentes genéticos de tipo salvaje, antecedentes genéticos con mutaciones de delección, antecedentes genéticos con otra integración diana y antecedentes genéticos con integraciones no diana.

- 25 El término "animal", como se usa en el presente documento, se refiere a un animal no humano. El animal puede ser un embrión, un juvenil o un adulto. Los animales adecuados incluyen vertebrados tales como mamíferos, aves, reptiles, anfibios, crustáceos y peces. Los ejemplos de mamíferos adecuados incluyen, sin limitarse a, roedores, animales de compañía, ganado y primates. Los ejemplos no limitantes de roedores incluyen ratones, ratas, hámsteres, jiribas y cobayas. Los animales de compañía adecuados incluyen, entre otros, gatos, perros, conejos, erizos y hurones. Los ejemplos no limitantes de ganado incluyen caballos, cabras, ovejas, cerdos, vacas, llamas y alpacas. Los primates adecuados incluyen, entre otros, monos capuchinos, chimpancés, lémures, macacos, titíes, tamarinos, monos araña, monos ardilla y monos de Vervet. Los ejemplos no limitantes de aves incluyen gallinas, pavos, patos y gansos. Como alternativa, el animal puede ser un invertebrado tal como un insecto, un nematodo, y similares. Los ejemplos no limitantes de insectos incluyen *Drosophila* y mosquitos. Un animal ejemplar es una rata. Los ejemplos no limitantes de cepas de ratas adecuadas incluyen las Dahl Salt-Sensitive, Fischer 344, Lewis, Long Evans Hooded, Sprague-Dawley y Wistar. En una realización, el animal no es un ratón genéticamente modificado. En cada una de las iteraciones anteriores de animales adecuados para la invención, el animal no incluye secuencias de transposones integradas al azar introducidas de forma exógena.
- 30

- 35 Un aspecto adicional de la presente divulgación proporciona células o líneas celulares genéticamente modificadas que comprenden al menos una secuencia cromosómica modificada. La célula o línea celular genéticamente modificada puede derivar de cualquiera de los animales genéticamente modificados desvelados en el presente documento. Como alternativa, la secuencia cromosómica se puede modificar en una célula como se describe anteriormente en el presente documento (en los párrafos que describen las modificaciones de la secuencia cromosómica en animales) usando los procedimientos descritos en el presente documento. La divulgación también engloba un lisado de dichas células o líneas celulares.
- 40

- 45 Las células son células eucariotas. Las células hospedadoras adecuadas incluyen hongos o levaduras, tales como *Pichia*, *Saccharomyces* o *Schizosaccharomyces*; células de insectos, tales como células SF9 de *Spodoptera frugiperda* o células S2 de *Drosophila melanogaster*; y células animales, tales como células de ratón, rata, hámster, primates no humanos, o humanos. Las células ejemplares son de mamífero. Las células de mamífero pueden ser células primarias. En general, se puede usar cualquier célula primaria que sea sensible a roturas bicitáreas. Las células pueden ser de una variedad de tipos celulares, por ejemplo, fibroblastos, mioblastos, linfocitos T o B, macrófagos, células epiteliales, y similares.
- 50

- 55 Cuando se usan líneas celulares de mamíferos, la línea celular puede ser cualquier línea celular establecida o una línea celular primaria que aún no esté descrita. La línea celular puede ser adherente o no adherente, o la línea celular puede crecer en condiciones que fomenten el crecimiento adherente, no adherente u organotípico utilizando técnicas estándar conocidas por personas expertas en la materia. En la sección (III) (g) del presente documento se proporcionan ejemplos no limitantes de células y líneas celulares de mamífero adecuadas. En otras realizaciones más, la célula puede ser una célula madre. En la sección (III) (g) se proporcionan ejemplos no limitantes de células madre adecuadas.

La presente divulgación también proporciona un embrión no humano modificado genéticamente que comprende al menos una secuencia cromosómica modificada. La secuencia cromosómica se puede modificar en un embrión como se describe anteriormente en el presente documento (en los párrafos que describen las modificaciones de la secuencia

cromosómica en animales) usando los procedimientos descritos en el presente documento. En una realización, el embrión es un embrión unicelular fertilizado no humano de la especie animal de interés. Los embriones de mamíferos ejemplares, incluyendo embriones unicelulares, incluyen sin limitarse a, ratón, rata, hámster, roedor, conejo, felino, canino, ovino, porcino, bovino, equino, y de primate.

5 **Definiciones**

A menos que se definan 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., Dictionary of Microbiology and Molecular Biology (2^a ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5^a Ed., R. Rieger y col. (eds.), Springer Verlag (1991); y Hale & Marham, The Harper Collins Dictionary of Biology (1991). Tal como se usa en el presente documento, los siguientes términos tienen los significados que se les asigna a menos que se especifique otra cosa.

10 Cuando se introducen elementos de la presente divulgación o la o las realizaciones preferidas de la misma, los artículos "un" o "una", "el", "la" y "dicho", "dicha" pretenden significar que hay uno o más de los elementos. Las expresiones "que comprende", "que incluye" y "que tiene" pretenden ser inclusivos y significan que puede haber elementos adicionales además de los elementos enumerados.

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

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

25 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 genético, estén o no adyacentes tales secuencias reguladoras a las secuencias de codificación y/o transcritas. Por consiguiente, un gen incluye, pero no necesariamente se limita a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada de ribosomas internos, potenciadores, silenciadores, aislantes, elementos límitrofes, orígenes de replicación, sitios de unión a matriz y regiones de control de locus.

30 El término "heterólogo" se refiere a una entidad que no es endógena ni nativa de la célula de interés. Por ejemplo, una proteína heteróloga se refiere a una proteína que deriva o derivó originalmente de una fuente exógena, tal como una secuencia de ácido nucleico introducida de forma exógena. En algunas ocasiones, la proteína heteróloga normalmente no es producida por la célula de interés.

35 Las expresiones "ácido nucleico" y "polinucleótido" se refieren a un desoxirribonucleótido o polímero de ribonucleótido, en conformación lineal o circular, y en forma monocatenaria o bicatenaria. Para los fines de la presente divulgación, estos términos no deben interpretarse como limitantes con respecto a la longitud de un polímero. Los términos pueden englobar análogos conocidos de nucleótidos naturales, así como los nucleótidos que están modificados en la base, restos del 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 se emparejará con T.

40 El término "nucleótido" se refiere a desoxirribonucleótidos o ribonucleótidos. Los nucleótidos pueden ser nucleótidos estándar (es decir, adenosina, guanosina, citidina, timidina y uridina) o análogos de nucleótidos. Un análogo de nucleótido se refiere a un nucleótido que tiene una base de purina o pirimidina modificada o un resto ribosa modificado. Un análogo de nucleótido puede ser un nucleótido de origen natural (por ejemplo, inosina) o un nucleótido de origen no natural. Los ejemplos no limitantes de modificaciones en los restos del azúcar o en la 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 con otros átomos (por ejemplo, 7-deaza purinas). Los análogos de nucleótidos también incluyen didesoxinucleótidos, 2'-O-metil nucleótidos, ácidos nucleicos bloqueados (LNA), ácidos péptido-nucleicos (PNA) y morfolinos.

Los términos "polipéptido" y "proteína" se usan indistintamente para referirse a un polímero de restos de aminoácidos.

50 Las técnicas para determinar la identidad de secuencias de aminoácidos y ácidos nucleicos se conocen en la técnica. Típicamente, 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, la identidad se refiere a una correspondencia exacta de nucleótido a nucleótido o de aminoácido a aminoácido de dos polinucleótidos o secuencias de polipéptidos, respectivamente. Se pueden comparar dos o más secuencias (polinucleótidos o aminoácidos) determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sean de ácido nucleico o secuencias 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 proporciona una alineación aproximada para las secuencias de ácido nucleico, Advances in Applied Mathematics 2:482-489 (1981). 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 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, y normalizada por Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). El Genetics Computer Group (Madison, Wisconsin) proporciona una implementación ejemplar de este algoritmo para determinar el porcentaje de identidad de una secuencia 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 alineación es BLAST, usado con parámetros por defecto. Por ejemplo, BLASTN y BLASTP se pueden usar con los siguientes parámetros predeterminados: código genético=estándar; filtro=nninguno; cadena=ambas; umbral=60; valores esperados=10; Matriz=BLOSUM62; Descripciones=50 secuencias; ordenar por= PUNTUACIÓN ALTA; Bases de datos=no redundantes, GenBank+EMBL+DDJB+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Los detalles de estos programas se pueden encontrar en el sitio web de GenBank.
- Como se podrían 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 dados a continuación, se interprete como ilustrativa y no en un sentido limitante.

Ejemplos

Los siguientes ejemplos ilustran ciertos aspectos de la invención.

20 Ejemplo 1: Modificación del gen Cas9 para expresión en mamífero

Un gen Cas9 de la cepa MGAS15252 de *Streptococcus pyogenes* (número de referencia YP_005388840.1) se optimizó con la preferencia de codones para *Homo sapiens* para mejorar su traducción en células de mamífero. El gen Cas9 también se modificó agregando una señal de localización nuclear PKKKRKV (SEQ ID NO:1) en el extremo C para dirigir a la proteína a los núcleos de las células de mamífero. La Tabla 1 muestra la secuencia de aminoácidos de Cas9 modificada, con la secuencia de localización nuclear subrayada. La Tabla 2 presenta la secuencia de ADN de Cas9 modificada con codones optimizados.

Tabla 1. Secuencia de aminoácidos de Cas9 modificada

```
MDKKYSIGLDIGTNSVGAWITDDYKVPSSKKFKVLGNTDRHSIKKNLIGALLFGSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEAMAKVDDSSFHRLEESFLVEEDKKHER
HPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFIEGDLN
PDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKR
NGLFGNLIALSLGLTPNFKNFDLAEDAKLQLSKDTYDDDDLNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNRKIEKILTFRIPYYGPLARGNSRFAWMTRK
SEETITPWNFEEVVDKGASAQSFIERNMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK
VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGAYHDLLKIJKDFLDNEENEDILEDIVLTTLFEDRGMIEERLKTYAHILFD
DKVMKQLKRRRTGWGRSLRKLINGIRDQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIIKKGILQTVKIVDELVKVMGHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQN
GRDMYVDQELDINRLSDYDWDHVQPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFQDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDLSRMTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVDVRKMIAKSEQUIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIWWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKKDLIILPKYSLFELENGRKMLASAGELQK
GNELALPSKYVNFLYASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLHQHSITGLYETRIDSQLGGDPKKKRKV (SEQ ID NO:9)
```

Tabla 2. Secuencia de ADN de Cas9 optimizada (5'-3')

```

ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTG
GGCGTGATCACCGACGACTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGG
GCAACACCGACCAGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTT
GGCTCTGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAA
GATACACCAAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACG
AGATGGCCAAGGTGGACGACAGCTTCCACAGACTGGAAGAGTCCTCCTGG
TGGAAAGAGGATAAGAACGACGAGCGGCACCCATCTCGGCAACATCGTGGAC
GAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAAGCTG
GCCGACAGCACCAGACAAGGCCGACCTGAGACTGATCTACCTGGCCCTGGCCCA
CATGATCAAGTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCGACAA
CAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGATCTACAATCAGCTGTT
CGAGGAAAACCCCATCAACGCCAGCAGAGTGGACGCCAAGGCCATCCTGAGCG
CCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCGGC
GAGAAGCGGAATGGCCTGTTGGCAACCTGATTGCCCTGAGCCTGGGCTGAC
CCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAG
CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACC
AGTACGCCGACCTGTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGA

```

(continuación)

Tabla 2. Secuencia de ADN de Cas9 optimizada (5'-3')

GCGACATCCTGAGAGTGAACAGCGAGATCACCAAGGGCCCCCTGTCCGCCTCT
 ATGATCAAGAGATACTGACGAGCACCACCAGGACCTGACCTGCTGAAAGCTCTC
 GTGCGGCAGCAGCTGCCCTGAGAAGTACAAAGAGATTTCCTCGACCAGAGCAAG
 AACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA
 GTTCATCAAGCCCCTGGAAAAGATGGACGGCACCAGGAAGTCTCGAA
 GCTGAACAGAGAGGACCTGCTGCCAACGGACTGCTCGACAACGGCAGCA
 TCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCCGGCAGGAA
 GATTTTACCCATTCCCTGAAGGACAACCAGGAAAAGATCGAGAAGATCCTGACCT
 TCAGAATCCCCACTACGTGGCCCTCTGCCAGGGAAACAGCAGATTGCC
 GGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTCGAGGAAGTG
 GTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTCGA
 TAAGAACCTGCCAACGAGAAAGGTGCTGCCAACGACAGCCTGCTGTACGAGTA
 CTTCACCGTGTACAACGAGCTGACCAAAGTGAATACGTGACCGAGGGAAATGCG
 GAAGCCCGCCTTCTGAGCGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGT
 TCAAGACCAACCAGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGA
 AAATCGAGTGCTCGACAGCGTGGAAATCAGCGCGTGGAAAGATCGTTCAACG
 CCTCCCTGGCGCCTATCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCC
 TGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGTGACCCGTGACAC
 TGTTGAGGACCGGGCATGATCGAGGAACGGCTGAAACCTATGCCACCTGT
 TCGACGACAAAGTGATGAAGCAGCTGAAGCAGCTGACCGGGATCCGGCTGGGC
 AGGCTGAGCCGGAAAGCTGATCAACGGCATCCGGACAAGCAGTCCGGCAAGAC
 AATCCTGGATTCCTGAAGTCCGACGGCTTCCGCAACAGAAACTTATGCAGCT
 GATCCACGACGACAGCCTGACCTTAAAGAGGACATCCAGAAAGCCCAGGTGTC
 CGGCCAGGGACACTCTGCACGAGCAGATGCCAATCTGCCGGATCCCCCG
 CCATTAAGAAGGGCATCCTGCAGACAGTGAAGATTGTTGACGAGCTCGTGAAG
 TGATGGGCCACAAGCCCAGAACATCGTGTGAAATGGCCAGAGAGAACCG
 ACCACCCAGAAGGGACAGAACAGCCCGAGAGAACAGCGGATCGAAGA
 GGGCATCAAAGAGCTGGGAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
 CCCAGCTGCAGAACGAGAACAGCTGTACCTGTACTACCTGCAGAACGGGATA
 TGTACGTGGACCAGGAACCTGGACATCAACCGGCTGCGACTACGATGTGGACC
 ACATTGTGCCCTAGTCCTCATCAAGGACGACTCCATCGATAACAAAGTGCTGAC
 TCGGAGCGACAAGAACCGGGCAAGAGCGACAACGTGCCCTCGAACAGAGGTC
 GTGAAGAAGATGAAGAAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACC
 CAGAGGAAGTTCGACAATCTGACCAAGGCCAGAGAGAGGCCCTGAGCGAACT
 GGATAAGGCCGGCTTCATTAAGCGGCAGCTGGTGGAAACCCGGCAGATCACAA
 AGCACGTGGCACAGATCCTGGACTCCGGATGAAACACTAAGTACGACGAGAAC
 GACAAACTGATCCGGGAAGTGAAGAGTGTACCCCTGAAAGTCCAAGCTGGTGTCC
 GACTTCAGAAAGGATTCCAGTTTACAAAGTGCAGAGATCAACAAACTACCACC
 ACGCCCACGACGCCTACCTGAACGCCGTGTTGGAACCGCCCTGATCAAAAG
 TACCTTAAGCTGGAAAGCGAGTTGTCGTGACGGCGATTACAAGGTGTACGACGTG
 CGGAAGATGATGCCAACAGAGCGAGCAGGAATCGGCAAGGCTACGCCAAGTA
 CTTCTTCTACAGCAACATCATGAACTTTCAAGACCGAGATCACACTGGCCAAC
 GGCGAGATCAGAAAGCGGCCCTGATCGAGACAAACGGCAGAACCGGGGAGAT
 CGTGTGGATAAGGGCCGGATTTCGCCACAGTGGAAAGTGCTGTCCATGC
 CCCAAGTGAATATCGTAAAAAGACCGAGGTGCAGACCGGGCTTCAGCAA
 GAGTCTATCCTGCCAACAGAGGAACCTCGACAGACTGATGCCAGAAAGAAGGAT

(continuación)

Tabla 2. Secuencia de ADN de Cas9 optimizada (5'-3')

```

TGGGACCCCTAAGAAGTACGGCGGCTTGACAGCCCCACCGTGGCCTACTCTGT
GCTGGTGGTGGCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGA
AAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTCGAGAAGAATCCA
TCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGGAAAAAGGACCTGATCATCA
AGCTGCCTAAGTACTCCCTGTTGAGCTGGAAAACGGCCGGAAGCAGGATGCTG
GCTTCTGCCGGCGAACTGCAGAAGGGAAACGAGCTGGCCCTGCCCTCCAAATA
TGTGAACCTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGA
GGATAATGAGCAGAAACAGCTGTTGGAACAGCACAAGCACTACCTGGACGA
GATCATCGAGCAGATTAGCGAGTTCTCCAAGCGCGTGATCCTGGCCGATGCCAA
CCTGGACAAGGTGCTGAGCGCCTACAACAAGCACCGGGATAAGCCCACATCAGAG
AGCAGGCCGAGAATATCATCCACCTGTTACCCCTGACCAACCTGGGAGGCCCTG
CCGCCTTCAAGTACTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCA
AAGAGGTGCTGGACGCCACCCCTGATCCACCAAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGAGGCACCCCAAGAAAAAGCGCAAAGT
G (SEQ ID NO:10)

```

La secuencia de ADN de Cas9 modificada se colocó bajo el control del promotor de citomegalovirus (CMV) para la expresión constituyente en células de mamífero. La secuencia de ADN de Cas9 modificada también se colocó bajo el control del promotor T7 para la síntesis de ARNm *in vitro* con la ARN polimerasa T7. La transcripción de ARN *in vitro* se realizó utilizando el kit de transcripción MessageMAX T7 ARCA-Capped Message y el sistema de producción de ARNm estándar T7 mScript (Cellscript).

5

Ejemplo 2: Direccionamiento de Cas9

El sitio de integración de virus adenoasociados 1 (AAVS1) se usó como una diana para la modificación del genoma humano mediada por Cas9. El locus AAVS1 humano se encuentra en el intrón 1 (4427 pb) de la proteína fosfatasa 1, subunidad reguladora 12C (PPP1R12C). La Tabla 3 muestra el primer exón (gris sombreado) y el primer intrón de PPP1R12C. La secuencia subrayada dentro del intrón es el sitio de modificación diana (es decir, el locus AAVS1).

10

Tabla 3. Primer exón e intrón de PPP1R12C (5'-3')

```

GC GG GCGGGCGGTGCGATGTC CGGAGAGGATGGCCCGGCGCTGGCCC GGG
GG CGGCGGGCGGCGGCTGCCC GGGAGCGGGAGCAGCTGCGGCAGTG
GGGGCGCGGGCGGGCGCCGAGCCTGGCCCCGGAGAGCAGCGCCGCGCCGCAC
CGTCCGCTTCGAGCGCGCCGCCGAGTTCTGGCGGCCGTGCGGGCGGCAC
CTGGACGAGGGCGTCTGATGCTGCGCGCCGCCGACCCCTGGCCCCGGCGCCG
AGCTCGACCCCGCCGCGCCGCCGCCCCGCGCCGTGCTGGACTCCACCAA

```

(continuación)

Tabla 3. Primer exón e intrón de PPP1R12C (5'-3')

CGCCGACGGTATCAGCGCCCTGCACCAGGTAGCGCCCCCGCCCCGGCGTCT
 CCCGGGGCCAGGTCCACCCCTTGCTGCCACCTGGGGCATCCTCCTCCCCG
 TTGCCAGTCTCGATCCGCCCCGTCGTTCTGCCCTGGGCTTGCCACCCATAG
 CTGACACCCCGTCCCAGTCCCCCTTACCAATTCCCCTCGACCACCCACTCCG
 AATTGGAGCCGCTTCAACTGGCCCTGGGCTTAGCCACTCTGTGCTGACCACCT
 GCCCCAGGCCTCCTTACCAATTCCCCTCGACCTACTCTTCCGATTGGAGTC
 GCTTAACTGGCCCTGGCTTGGCAGGCTGTGCTGACCCATGCAGTCCTCCTTA
 CCATCCCTCCCTCGACTTCCCCTTCCGATGTTGAGGCCCTCCAGCCGGTCCT
 GGACTTTGTCTCCCTGCCCTGCCCTCTCCTGAACCTGAGCCAGCTCCCAT
 AGCTCAGTCTGGTCTATCTGCCCTGGCCCTGGCCATTGTCACTTGCCTGCCCT
 CCTCTCGCCCCGAGTGCCCTGCTGTGCCGCCGAACCTGCCCTTAACGCT
 GCCGTCTCTCCTGAGTCCGGACCACTTGAGCTCTACTGGCTCTGCGCCGC
 CTCTGGCCCACGTGTTCCCCTCCAGGCAGGTCTGCTTGAGCTGGCCTGGTCACCT
 CTACGGCTGGCCCAGATCCTCCCTGCCGCCCTTCAGGTTCCGTCTCCTCC
 ACTCCCTCTTCCCCTGCTCTGCTGTGTTGCCAAGGATGCTCTTCCGG
 AGCACTTCCTCTCGCGCTGCACACAGTGTGCTGCCAAGGATGCTCTTCCGG
 GTGTCTGGGTCCCTCCGGCATCTCCCTCCCTCACCCAACCCATGCCGTCT
 TCACTCGCTGGGTTCCCTTTCTTCTGGGGCTGTGCCATCTCGT
 TCTTAGGATGGCCTCTCCGACGGATGTCCTCCCTGCGTCCCGCTCCCTCT
 TGTAGGCCTGCATCATCACCGTTCTGGACAAACCCAAAGTACCCCGTCTCCC
 TGGCTTAGGCACCTCTCCATCCTCTTGCCTTGACACCCCGTTCTC
 CTGTGGATTGGGTACCTCTCACTCCTTCATTGGCAGCTCCCTACCC
 CTTACCTCTAGTCTGTGCTAGCTCTCCAGCCCCCTGTATGGCATCTCCAG
 GGGTCCGAGAGCTCAGCTAGTCTTCTCCCAACCCGGCCCTATGTCCACT
 TCAGGACAGCATGTTGCTGCCCTCAGGGATCCTGTGTCGGAGCTGGGACCA
 CCTTATATTCCCAGGGCCGGTTAATGTGGCTGGTTGGTACTTTATCTGT
CCCCTCCACCCACAGTGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCC
 CCATCCTTAGGCCTCCCTCCTAGTCTCTGATATTGGCTAACCCCCACCT
 CCTGTTAGGCAGATTCTTATCTGGTGACACACCCCCATTCCCTGGAGCCATCTC
 TCTCCTGCCAGAACCTCTAAGGTTGCTACGATGGAGCCAGAGAGGATCCTG
 GGAGGGAGAGCTGGCAGGGGGTGGAGGGAAAGGGGGGATGCGTGACCTG
 CCCGGTTCTCAGTGGCCACCCCTGCGCTACCCCTCCCAGAACCTGAGCTGCTCT
 GACGCGGCCGTCTGGTGCCTTCACTGATCCTGGTGCAGCTCCTTACACT
 TCCCAAGAGGAGAAGCAGTTGGAAAAACAAATCAGAATAAGTGGCCTGAG
 TTCTAACCTTGCTCTCACCTTCTAGTCCCCAATTATATTGGCTCCGTGCG
 TCAGTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCTGGCAGGGCTGT
 GGTGAGGAGGGGGGTGTCGTGGAAAACCTCCCTTGTGAGAATGGTGCCTC
 CTAGGTGTTCACCAAGGTCGTGGCCGCCTCTACTCCCTTCTCTTCTCCATCCT
 CTTCCCTAAAGAGTCCCCAGTGCTATCTGGACATATTCCCTCCGCCAGAGCA
 GGGTCCCGCTCCCTAAGGCCCTGCTCTGGCTTCTGGTTGAGTCCTGGCA
 AGCCCAGGAGAGGCCAGGCCACCTACTTGGCCTGGACCCACGAGAGGCCACCC
 TGCCCCCTGGCTCTCCTGCCCTCCCTACAGGGGTTCTGGCTCTGCTCTCAG
 ACTGAGCCCCGTTCCCTGCATCCCCGTTCCCTGCATCCCCCTTCCCTGCAT
 CCCCCAGAGGCCAGGCCACCTACTTGGCCTGGACCCACGAGAGGCCACCC
 CAGCCCTGTCTACCAGGCTGCCTTGGTGATTCTCCTCCAACTGTGGGGTG

(continuación)

Tabla 3. Primer exón e intrón de PPP1R12C (5'-3')

```

ACTGCTTGGCAAACACTCACTCTCGGGGTATCCCAGGAGGCCTGGAGCATTGGG
GTGGGCTGGGTTCAGAGAGGGAGGATTCCCTCTCAGGTTACGTGGCCAAGA
AGCAGGGGAGCTGGTTGGTCAGGTCTGGGTGTGGGTGACCAGCTTATGC
TGTGCCCCAGGACAGCCTAGTTAGCACTGAAACCCCTCAGTCCTAGGAAAACA
GGGATGGTTGGCACTGTCTGGGTGACTCTGATTCCCGGCCAGTTCTCCA
CCTGGGGCTGTGTTCTCGTCCTGCATCCTCTCCAGGCAGGTCCCCAACGCATC
GCCCGCCCTGCTGTGGCTGTTCCAAGTTCTTAGGGTACCCCACGTGGGTATC
AACCACTTGGTGAGGCTGGTACCCCTGCCCTTCAGGTCCCCAACGCCTTA
GTGGCTAGGGGGTTGGGGCTAGAGTAGGAGGGCTGGAGGCCAGGATTCTTAG
GGCTGAACAGAGAAGAGCTGGGGCCTGGGCTCCTGGGTTGAGAGAGGAGG
GGCTGGGGCCTGGACTCCTGGGTCCAGGGAGGGCTGGACTCCTGGTCCGAGGGA
CCTGGGTCTGAGGGTGGAGGGACTGGGGCCTGGACTCCTGGTCCGAGGGA
GGAGGGGCTGGGCCTGGACTCGTGGTCTGAGGGAGGAGGGCTGGGCTGGACT
CTGGACTTCTGGGTCTTAGGGAGGCCTGGACTCCTCATCTGAGGGCGGAAGGGCTGG
ATGGGGAGAGGCTGGGGCCTGGACTCCTCATCTGAGGGCGGAAGGGCTGG
GCCCTGGCCTCCTGGTTGAATGGGGAGGGTGGGCCTGGACTCTGGAGTCC
CTGGTGCCCAAGGCCTCAGGCATCTTACAGGGATGCCTGTACTGGCAGGTC
CTTGAAGGGAAAGGCCATTGCTCTCCTGCCCTCCCTATGCCATGAC
AACTGGGTGGAAATAAACGAGGCCAGTTCCATGACCTCATGCTCTGGCCCTCGTAGCTCC
CCCCTCACAGCCCAGTTCCATGACCTCATGCTCTGGCCCTCGTAGCTCC
TCCCGCCTCCTCCAGATGGCAGCTTGGAGAGGTGAGGGACTGGGGGTAA
TTTATCCCCTGGATCTAGGAGTTAGCTTCACTCCTCCTCAGCTCCAGTTCAAGG
TCCCGGAGCCCACCCAGTGTCCACAAGGCCTGGGCAAGTCCCTCCGACC
CCCTGGACTTGGCTTTGTCCCCCAAGTTTGGACCCCTAAGGGAAAGAATGA
GAAACGGTGGCCGTGTCAAGCCCTGGCTGCAGGGCCCCGTGCAGAGGGG
CTCAGTGAAGTGGAGTGTGACAGCCTGGGCCCCAGGCACACAGGTGTGCAGCT
GTCTCACCCCTCTGGAGTCCCCTGGCCAGGCCCCCTGAGTCTGTCCCAGCACAGG
GTGGCCTCCTCCACCCCTGCATAGCCCTGGGCCACGGCTCGTCCCTGCAGA
GTATCTGCTGGGTGGTTCCGAGCTTGAACCCCTGGAAAGGACCTGGCTGGGTT
AAGGCAGGAGGGGCTGGGGCCAGGACTCCTGGCTCTGAAGGGAGGAGGGCT
GGAACCTCTCCCTAGTCTGAGCACTGGAAAGGCCACCTGTGGGTGGTAGCGG
GGGTTTGCCTGTCTAACAGGTACCATGTGGGTTCCGCACCCAGATGAGAA
GCCCTCCCTCCCTCCCTGCAGATAGCCAGGAGTCCTTCTGGCTGGGTGGTAGCGG
GGTTTCACTGAGCACTGAAGGCCCTGGCCGGCCTGACCAACTGGCAACCAGGC
GTATCTAAACAGCCAGTGGCCAGAGGCTGTGGGTCTTCCCTGGAGTCCCAGCTGCTA
GCACCGTGTCCCTGGATCTGTTCTGGCTCCCTCTGGAGTCCCAGCTGCTG
GGACACCGTGGCTGGGTAGGTGCGGCTGACGGCTGTTCCCACCCCCAG
(SEQ ID NO:11)

```

Se diseñaron ARN guía de Cas9 para dirigirse al locus AAVS1 humano. Se prepararon un ARN de 42 nucleótidos (denominado en el presente documento secuencia de "ARNcr") que comprende (5' a 3') una secuencia de reconocimiento de la diana (es decir, secuencia complementaria a la cadena no codificante de la secuencia diana) y una secuencia protoespaciadora; un ARN de 85 nucleótidos (denominado en el presente documento secuencia de "ARNtracr") que comprende la secuencia 5' con complementariedad con la secuencia 3' del ARNcr y con la secuencia horquilla adicional; y un ARN químérico que comprende los nucleótidos 1-32 del ARNcr, un bucle GAAA y los nucleótidos 19-45 del ARNtracr. El ARNcr se sintetizó químicamente por Sigma-Aldrich. El ARNtracr y el ARN químérico se sintetizaron por transcripción *in vitro* con la ARN polimerasa T7 utilizando el kit T7-Scribe Standard RNA IVT (Cellscript). La secuencia codificante de ARN químérico también se colocó bajo el control del promotor U6 humano para la transcripción *in vivo* en células humanas. La Tabla 4 muestra las secuencias de los ARN guía.

5

10

Tabla 4. ARN guía

| ARN | Secuencia 5'-3' | SEQ ID NO: |
|---------------|------------------------------------------------------------------------------------|------------|
| AAVS1-ARNcr | ACCCACAGUGGGGCCACUAGUUUUAGAGCUAUGCUGUUUG | 12 |
| ARNtracr | GGAACCAIUCAAAACAGCAUAGCAAGUUAAAUAAGGCUAGUCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUCUUUUUU | 13 |
| ARN químérico | ACCCACAGUGGGGCCACUAGUUUUAGAGCUAGAAAUAAGGCUAGUCG | 14 |

Ejemplo 3: Preparación del polinucleótido donador para controlar la modificación del genoma

Se usó la integración dirigida de una proteína GFP en el extremo N de PPP1R12C para controlar la modificación del genoma mediada por Cas9. Para mediar la integración por recombinación homóloga se preparó un polinucleótido donador. El donador de ADN AAVS1-GFP contenía un brazo homólogo 5' del locus AAVS1 (1185 pb), un receptor de empalme de ARN, una secuencia de codificación turbo GFP, un terminador de transcripción 3' y un brazo homólogo 3' del locus AAVS1 (1217 pb). La Tabla 5 muestra las secuencias del receptor de empalme de ARN y la secuencia de codificación GFP seguida del terminador de transcripción 3'. El ADN plasmídico se preparó utilizando el kit GenElute Endotoxin-Free Plasmid Maxiprep (Sigma).

Tabla 5. Secuencias en la secuencia donadora de ADN AAVS1-GFP

| | Secuencia 5'-3' | SEQ ID NO: |
|-------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Receptor de empalme de ARN | CTGACCTCTTCTCTCCCACAG | 15 |
| Secuencia de codificación GFP y terminador de transcripción | GCCACCATGGACTACAAAGACGATGACGACAAGGTCGACTCTAGAGCTGCAGAGAGCGACGAGAGCGGCCTGCCGCCATGGAGATCGAGTGCCGCATCACCGGCACCCCTGAACGGCGTGGAGTTCGAGCTGGTGGCGGGAGAGGGCACCCCCAGCAGGGCGAGCAGGGCCCGCATGACCAACAAGATGAAGAGCACCAAAGGCCCTGACCTTCAGCCCCAACCTGCTGAGCCACGTGATGGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCTACGAGAACCCCTTCCTGCACGCCATCAACAAACGGCGGCTCACCAACACCCGCATCGAGAAAGTACGAGGACGGCGGGCTGCTGCACGTGAGCTTCAGCTACCGCTACGAGGGCCGGCGCGTGTACGGCGACTTCAAGGTGATGGGCACCGGCTTCCCCGAGGACAGCGTGATCTCACCGACAAGATCGTCCGCAGCACGCCACCGTGGAGCACCTGCACCCATGGCGATAACGATCTGGATGGCAGCTTCACCCGCACCTTCAGCCTGCGCGACGGCGCTACTACAGCTCCGTGGTGGACAGCCACATGCACTTCAAGAGCGCCATCCACCCAGCATCCTGCAGAACGGGGGCCCATGTTGCCTCCCGCCGCGTGGAGGGAGGATCACAAGAACCCGGATGCAGATGCCGGTAAGAATGAAGATCTCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCGTGCCTTCCCTGACCCCTGGAAAGGTGCCACTCCACTGTCCTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGCAGGACAGCAAGGGGAGGATTGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGCTATGGACTCGAGGTTAACGTCGACGCGGCCGCGT | 16 |

10 La integración genética dirigida dará como resultado una proteína de fusión entre los primeros 107 aminoácidos del PPP1R12C y el turbo GFP. La proteína de fusión esperada contiene los primeros 107 restos de aminoácidos de

PPP1R12C (resaltados en gris) del empalme de ARN entre el primer exón de PPP1R12C y el receptor de empalme diseñado por ingeniería genética (véase Tabla 6).

Tabla 6. Secuencia de aminoácidos prevista de la proteína de fusión PPP1R12C-GFP.

| |
|----------------------------------------------------------------|
| MSGEDGPAAGPGAAAAAARERRREQLRQW GARAGAE PGPGERRARTVRFERA AE |
| FLAACAGGDLDEARLMLRAADPGPGAELDPAAPPPARAVLDSTNADGISALHQATM |
| DYKDDDDKVDSRAAESDESGLPAMEIECRITGTLNGVEFELVGGEGTPEQGRMTN |
| KMKSTKGALTFS PYLLSHVMGYGFYHFGTYP SGYENPFLHAINNGGYTNTRIEKYED |
| GGVLHVSFSYRYEAGR VIGDFKVMGTGF PEDSVIFTDKIVRSNATVEHLHPMGDNDL |
| DGSFTRTFSLRDGGYYSSVV DSHMF KSAIHPSI LQNGGP MFA FRR VEEDHS NTEL |
| GIVEYQHAFKTPDADAGEE (SEQ ID NO:17) |

Ejemplo 4: Integración dirigida mediada por Cas9

- 5 La transfección se realizó en células K562 humanas. La línea celular K562 se obtuvo de la American Type Culture Collection (ATCC) y se cultivó en medio de Dulbecco modificado por Iscove, suplementado con 10 % de FBS y L-glutamina 2 mM. Todos los medios y suplementos se obtuvieron de Sigma-Aldrich. Los cultivos se dividieron un día antes de la transfección (aproximadamente 0,5 millones de células por ml antes de la transfección). Las células se transfecaron con Nucleofector Solution V (Lonza) en un Nucleofector (Lonza) con el programa T-016. Cada nucleofección contenía aproximadamente 0,6 millones de células. Los tratamientos de transfección se detallan en la Tabla 7. Las células se cultivaron a 37 °C y 5 % de CO₂ inmediatamente después de la nucleofección.
- 10

Tabla 7. Tratamientos de transfección.

| Tratamiento | Cas9 modificado | ARN guía | Secuencia donadora |
|-------------|--------------------------------------------------------------------------------------------|-----------------------------------------------|-------------------------------------|
| A | ARNm de Cas9 transcripto con un análogo de caperuza anti-inverso (10 µg) | dúplex pre-anillado ARNcr-ARNtracr (0,3 nmol) | ADN plasmídico de AAVS1-GFP (10 µg) |
| B | ARNm de Cas9 transcripto con un análogo de caperuza anti-inverso (10 µg) | ARN químérico (0,3 nmol) | ADN plasmídico de AAVS1-GFP (10 µg) |
| C | ARNm de Cas9 encapuchado a través de la reacción de encapuchado postranscripcional (10 µg) | ARN químérico (0,3 nmol) | ADN plasmídico de AAVS1-GFP (10 µg) |
| D | ADN plasmídico de Cas9 (10 µg) | ADN plasmídico de ARN químérico de U6 (5 µg) | ADN plasmídico de AAVS1-GFP (10 µg) |
| E | Nada | Nada | ADN plasmídico de AAVS1-GFP (10 µg) |
| F | Nada | Nada | Nada |

La clasificación de células activadas por fluorescencia (FACS) se realizó 4 días después de la transfección. Los datos de FACS se muestran en la **Figura 2**. El porcentaje de GFP detectado en cada uno de los cuatro tratamientos experimentales (A-D) fue mayor que en los tratamientos control (E, F), confirmando la integración de la secuencia donadora y la expresión de la proteína de fusión.

15 **Ejemplo 5: Confirmación por PCR de integración dirigida**

El ADN genómico se extrajo de las células transfecadas con el kit GenElute Mammalian Genomic DNA Miniprep (Sigma) 12 días después de la transfección. Luego, el ADN genómico se amplificó por PCR con un cebador directo ubicado fuera del brazo homólogo 5' del plásmido donador AAVS1-GFP y un cebador inverso ubicado en la región 5' de la GFP. El cebador directo fue 5'-CCACTCTGTGCTGACCACTCT-3' (SEQ ID NO: 18) y el cebador inverso fue 5'-GC GGCACTCGATCTCCA-3' (SEQ ID NO: 19). El tamaño de fragmento esperado de la PCR de unión fue de 1388 pb. La amplificación se realizó con JumpStart Taq ReadyMix (Sigma), utilizando las siguientes condiciones de ciclo: 98 °C durante 2 minutos para la desnaturización inicial; 35 ciclos de 98 °C durante 15 segundos, 62 °C durante 30 segundos y 72 °C durante 1 minuto y 30 segundos; y una extensión final a 72 °C durante 5 minutos. Los productos de PCR se resolvieron en gel de agarosa al 1%.

Células transfectadas con 10 µg de ARNm de Cas9 transcritas con un análogo de caperuza anti-inverso, 0,3 nmol del dúplex ARNcr-ARNtracr pre-anillado, y 10 µg de ADN plasmídico de AAVS1-GFP exhibieron un producto de PCR del tamaño esperado (véase el carril A, **Figura 3**).

Ejemplo 6: Edición del genoma basada en Cas9 en embriones de ratón

- 5 El locus Rosa26 de ratón puede ser objeto de modificaciones genómicas. La Tabla 8 presenta una parte de la secuencia de Rosa26 del ratón en la que los sitios diana potenciales se muestran en negrita. Cada sitio diana comprende un protoespaciador.

Tabla 8. Secuencia de Rosa26 de ratón

| |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GAGCGGCTGCGGGCGGGTGCAAGCACGTTCCGACTTGAGTTGCCTCAAGAG GGCGTGCTGAGCCAGACCTCCATCGCGCACTCCGGGGAGTGGAGGGAGGAAGGA GCGAGGGCTCAGTTGGGCTGTTTGGAGGCAGGAAGCAGTGCTCTCCCAAAGT CGCTCTGAGTTGTTATCAGTAAGGGAGCTGCAGTGGAGTAGGCAGGGAGAAGG CCGCACCCTCTCCGGAGGGGGAGGGAGTGTGCAATACCTTCTGGGAGT TCTCTGCTGCCTCCTGGCTCTGAGGACCGCCCTGGGCCTGGAGAATCCCTTC CCCCTCTCCCTCGTGTGCAACTCCAGTCTTCTAGAAGATGGCGGGAGT CTTCTGGGCAGGCTAAAGGCTAACCTGGTGTGGCGTTGTCCTGCAGGGG AATTGAACAGGGTGTAAAATTGGAGGGACAAGACTTCCCACAGATTTCGGTTT GTCGGGAAGTTTTAATAGGGCAAATAAGGAAATGGGAGGATAGGTAGTCA TCTGGGGTTTATGCAGAAAACACAGGTTATTATTGCTGTGATCCGCCTCGG AGTATTTCATCGAGGTAGATTAAAGACATGCTCACCGAGTTTACTCTCCT GCTTGAGATCCTTAACAGTATGAAATTACAGTGTGCGAGTTAGACTATGTAA GCAGAATTAA (SEQ ID NO:20) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Se diseñaron ARN guía para dirigirse a cada uno de los sitios diana en el locus Rosa26 de ratón. Las secuencias se muestran en la Tabla 9, cada uno tiene 42 nucleótidos de longitud y la región 5' es complementaria a la cadena que no se presenta en la Tabla 8 (es decir, la cadena que es complementaria a la cadena que se muestra en la Tabla 8).

Tabla 9. ARN guía de Rosa26 de ratón

| ARN | Secuencia 5'-3' | SEQ ID NO: |
|-----------------|------------------------------------------------|------------|
| mRosa26-ARNcr-1 | CUCCAGUCUUUCUAGAAGAUGUUUUAGAGCUAU GCUGUUUUG | 21 |
| mRosa26-ARNcr-2 | UGAACAGGUGUAAAAUUGGAGUUUUAGAGCUAU GCUGUUUUG | 22 |
| mRosa26-ARNcr-3 | UGUCGGGAAGUUUUUUAAUAGUUUUAGAGCUAU GCUGUUUUG | 23 |

Los ARNcr se sintetizaron químicamente y se pre-anillaron al ARNtracr (SEQ ID NO: 13; véase el ejemplo 2). El ARNcr/ARNtracr y el ARNm transrito *in vitro* que codifica la proteína Cas9 modificada (SEQ ID NO: 9; véase el ejemplo 1) puede microinyectarse en los pronúcleos de embriones de ratón fertilizados. Bajo la guía a la diana establecida por el ARNcr, la proteína Cas9 escinde el sitio diana, y la ruptura bicatenaria resultante se puede reparar mediante un proceso de reparación de unión de extremos no homólogos (NHEJ). Los embriones inyectados pueden incubarse a 37 °C, 5 % de CO₂ durante la noche o durante hasta 4 días, seguido de un análisis de genotipado, o los embriones inyectados pueden implantarse en ratones hembra receptores de modo que los animales nacidos vivos puedan genotiparse. En los embriones o tejidos incubados *in vitro* de animales nacidos vivos puede explorarse la presencia de mutación inducida por Cas9 en el locus Rosa utilizando procedimientos estándar. Por ejemplo, los embriones o tejidos del feto o de animales vivos se pueden recolectar para la extracción y análisis del ADN. El ADN puede aislar utilizando procedimientos estándar. La región diana del locus Rosa26 puede amplificarse mediante PCR usando cebadores apropiados. Debido a que el NHEJ es propenso a errores, pueden producirse delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la ruptura. Las mutaciones se pueden detectar utilizando procedimientos de genotipado basados en PCR, tales como ensayos de emparejamiento erróneo Cel-I y secuenciación de ADN.

Ejemplo 7: Modificación del genoma basada en Cas9 en embriones de ratón

El locus Rosa26 se puede modificar en embriones de ratón mediante la inyección conjunta de un polinucleótido donador, como se detalla anteriormente en la sección (III) (d), junto con el ARNcr/ARNtracr pre-anillado y el ARNm que codifica Cas9 modificado como se describe anteriormente en el ejemplo 6. En los embrones o tejidos incubados *in vitro* de animales vivos (como se describe en el ejemplo 6) puede explorarse la presencia de un locus Rosa26 modificado utilizando procedimientos de genotipado basados en PCR, tales como ensayos RFLP, PCR de uniones y secuenciación de ADN.

5

Ejemplo 8: Edición del genoma basada en Cas9 en embriones de rata

El locus Rosa26 de rata puede ser objeto de modificaciones genómicas. La Tabla 10 muestra una parte de la secuencia de rata en la que los sitios diana potenciales se muestran en negrita. Cada sitio diana comprende un protoespaciador.

Tabla 10. Secuencia de Rosa26 de rata

```

GGGATTCCCTCCTTGAGTTGTGGCACTGAGGAACGTGCTGAACAAGACCTACATT
GCACTCCAGGGAGTGGATGAAGGAGTTGGGCTCAGTCGGTTGTATTGGAGA
CAAGAACACTTGCTCTCCAAAAGTCGGTTGAGTTATCATTAAAGGGAGCTGCAG
TGGAGTAGGCGGAGAAAAGGCCGCACCCCTCTCAGGACGGGGGAGGGGAGTG
TTGCAATACTTCTGGAGTTCTGCTGCCTCCTGTCTTGAGGACCGCCCT
GGGCCTGGAAGATTCCCTCCCCCTTCCCTCGTATCTGCAACTGGAGTCT
TTCTGGAAGATAGGCGGGAGTCTTCTGGGCAGGCTAAAGGCTAACCTGGTGC
GTGGGGCGTTGTCCTGCAGAGGAATTGAACAGGTGAAAATTGGAGGGGCAAG
ACTTCCCACAGATTTGATTGTGTTAAGTATTGTAATAGGGGCAAATAAGG
GAAATAGACTAGGCACTCACCTGGGGTTTATGCAGCAAAACTACAGGTTATTAT
TGCTTGTGATCCGCCCTGGAGAATTTCACCGAGGTAGATTGAAGACATGCC
ACCCAAATTAAATATTCTTCACTTGCACCTGCGATCCTGCTACAGTATGAAA (SEQ ID
NO:24)

```

- 10 Se diseñaron ARN guía para dirigirse a cada uno de los sitios diana en el locus Rosa26 de rata. Las secuencias se muestran en la Tabla 11, cada uno tiene 42 nucleótidos de longitud y la región 5' es complementaria a la cadena que no se presenta en la Tabla 10 (es decir, la cadena que es complementaria a la cadena que se muestra en la Tabla 10).

Tabla 11. ARN guía de Rosa26 de rata

| ARN | Secuencia 5'-3' | SEQ ID NO: |
|-----------------|------------------------------------------------|------------|
| rRosa26-ARNcr-1 | AGGGGGAAGGGAAUCUUCAGUUUAGAGCUA UGCUGUUUUG | 25 |
| rRosa26-ARNcr-2 | UCUGCAACUGGAGUCUUUCUGUUUAGAGCUA UGCUGUUUUG | 26 |
| rRosa26-ARNcr-3 | AGGCAGGGAGUCUUCUGGGCAGUUUAGAGCUA UGCUGUUUUG | 27 |

- 15 Los ARNcr se sintetizaron químicamente y se pre-anillaron al ARNtracr (SEQ ID NO: 13; véase el ejemplo 2). El ARNcr/ARNtracr y el ARNm transcrita *in vitro* que codifica la proteína Cas9 modificada (SEQ ID NO: 9; véase el ejemplo 1) puede microinyectarse en los pronúcleos de embriones de rata fertilizados. Tras la guía hacia el sitio diana por parte del ARNcr, la proteína Cas9 escinde el sitio diana, y la ruptura bicatenaria resultante se puede reparar mediante un proceso de reparación de extremos no homólogos (NHEJ). Los embriones inyectados pueden incubarse a 37 °C, 5 % de CO₂ durante la noche o durante hasta 4 días, seguido de un análisis de genotipado, o los embriones inyectados pueden implantarse en ratones hembra receptores de modo que los animales nacidos vivos puedan genotiparse. En los embriones o tejidos incubados *in vitro* de animales nacidos vivos puede explorarse la presencia de mutación inducida por Cas9 en el locus Rosa utilizando procedimientos estándar. Por ejemplo, los embriones o tejidos del feto o de animales vivos se pueden recolectar para la extracción y análisis del ADN. El ADN puede aislarse utilizando procedimientos estándar. La región diana del locus Rosa26 puede amplificarse mediante PCR usando cebadores apropiados. Debido a que el NHEJ es propenso a errores, pueden producirse delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas durante la reparación de la ruptura. Las mutaciones se pueden detectar utilizando procedimientos de genotipado basados en PCR, tales como ensayos de emparejamiento erróneo Cel-I y secuenciación de ADN.
- 20
- 25

Ejemplo 9: Modificación del genoma basada en Cas9 en embriones de rata

El locus Rosa26 se puede modificar en embriones de rata mediante la inyección conjunta de un polinucleótido donador, como se detalla anteriormente en la sección (III) (d), junto con el ARNcr/ARNtracr pre-anillado y el ARNm que codifica Cas9 modificado como se describe anteriormente en el ejemplo 8. En los embriones o tejidos incubados *in vitro* de ratas nacidas vivas (como se describe en el ejemplo 8) puede explorarse la presencia de un locus Rosa26 modificado utilizando procedimientos de genotipado basados en PCR, tales como ensayos RFLP, PCR de uniones y secuenciación de ADN.

LISTADO DE SECUENCIAS

<110> SIGMA-ALDRICH CO. LLC

CHEN, Fuqiang

10 DAVIS, Gregory D.

<120> MODIFICACIÓN Y REGULACIÓN DEL GENOMA BASADA EN CRISPR

<130> P2502EP07

<140> PCT/US2013/073307

<141> 2013-12-05

15 <150> US 61/734.256
<151> 06/12/2012

<150> US 61/758.624
<151> 30/01/2013

20 <150> US 61/761.046
<151> 05/02/2013

<150> US 61/794.422
<151> 15/03/2013

<160> 27

<170> PatentIn versión 3.5

25 <210> 1
<211> 7
<212> PRT
<213> Secuencia artificial

<220>

30 <223> SINTETIZADO

<400> 1

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| Pro | Lys | Lys | Lys | Arg | Lys | Val |
| 1 | | | | | 5 | |

<210> 2

<211> 7

35 <212> PRT
<213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 2

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| Pro | Lys | Lys | Lys | Arg | Arg | Val |
| 1 | | | | 5 | | |

<210> 3

<211> 16

<212> PRT

<213> Secuencia artificial

45 <220>

<223> SINTETIZADO

ES 2 757 808 T3

<400> 3

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

<210> 4

<211> 20

5 <212> PRT

<213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 4

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

10 Lys Arg Lys Val
20

<210> 5

<211> 19

<212> PRT

<213> Secuencia artificial

15 <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

20 Arg Lys Val

<210> 6

<211> 24

<212> PRT

<213> Secuencia artificial

<220>

<223> SINTETIZADO

25 <400> 6

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

20 Ala Pro Lys Lys Arg Lys Val
20

<210> 7

<211> 27

<212> PRT

30 <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

<210> 8

<211> 21

<212> PRT

5 <213> Secuencia artificial

<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

10 <210> 9

<211> 1374

<212> PRT

<213> Secuencia artificial

<220>

15 <223> SINTETIZADO

<400> 9

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80

ES 2 757 808 T3

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

ES 2 757 808 T3

325

330

335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

ES 2 757 808 T3

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830

ES 2 757 808 T3

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

ES 2 757 808 T3

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

ES 2 757 808 T3

1310

1315

1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
 1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
 1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Pro
 1355 1360 1365

Lys Lys Lys Arg Lys Val
 1370

<210> 10

<211> 4122

<212> ADN

5 <213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 10

| | | | | | | |
|------------|------------|-------------|-------------|------------|--------------|------|
| atggacaaga | agtacagcat | cggcctggac | atcggcacca | actctgtggg | ctggccgtg | 60 |
| atcaccgacg | actacaagg | gcccagcaag | aaattcaagg | tgctggcaa | caccgaccgg | 120 |
| cacagcatca | agaagaacct | gatcggcgcc | ctgctgttcg | gctctggcga | aacagccgag | 180 |
| gccacccggc | tgaagagaac | cgccagaaga | agatacacca | gacggaagaa | ccggatctgc | 240 |
| tatctgcaag | agatcttcag | caacgagatg | gccaagggtgg | acgacagctt | cttccacaga | 300 |
| ctggaagagt | ccttcctgg | ggaagaggat | aagaagcacf | agcggcaccc | catcttcggc | 360 |
| aacatcgtgg | acgaggtggc | ctaccacgag | aagtacccca | ccatctacca | cctgagaaaag | 420 |
| aagctggccg | acagcaccga | caaggccgac | ctgagactga | tctacctggc | cctggcccac | 480 |
| atgatcaagt | tccggggcca | cttcctgatc | gagggcgacc | tgaacccca | caacagcgac | 540 |
| gtggacaagc | tgttcatcca | gctggcgcag | atctacaatc | agctttcga | ggaaaacccc | 600 |
| atcaacgcca | gcagagtgg | cgccaaggcc | atcctgagcg | ccagactgag | caagagcaga | 660 |
| cggctggaaa | atctgatcgc | ccagctgccc | ggcgagaagc | ggaatggcct | gttcggcaac | 720 |
| ctgattgccc | tgagcctgg | cctgacccccc | aacttcaaga | gcaacttcga | cctggccgag | 780 |
| gatgccaaac | tgcagctgag | caaggacacc | tacgacgacg | acctggacaa | cctgctggcc | 840 |
| cagatcggcg | accagtacgc | cgacctgttt | ctggccgcca | agaacctgtc | cgacgcccattc | 900 |
| ctgctgagcg | acatcctgag | agtgaacagc | gagatcacca | aggccccct | gtccgcctct | 960 |
| atgatcaaga | gatacgacga | gcaccaccag | gacctgaccc | tgctgaaagc | tctcgtgcgg | 1020 |
| cagcagctgc | ctgagaagta | caaagagatt | ttcttcgacc | agagcaagaa | cggtacgccc | 1080 |

| | |
|--------------------------------------------------------------------|------|
| ggctacatcg atggcggagc cagccaggaa gagttctaca agttcatcaa gcccattctg | 1140 |
| gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg | 1200 |
| aagcagcggc ccttcgacaa cgccagcatc cccaccaga tccaccctggg agagctgcac | 1260 |
| gccattctgc ggccggcagga agattttac ccattcctga aggacaaccg ggaaaagatc | 1320 |
| gagaagatcc tgaccttcag aatcccctac tacgtgggcc ctctggccag gggaaacagc | 1380 |
| agattcgcct ggatgaccag aaagagcgag gaaaccatca cccccctggaa cttcgaggaa | 1440 |
| gtggtggaca agggcgccag cgccccagagc ttcatcgagc ggatgaccaa cttcgataag | 1500 |
| aacctgccc acgagaaggt gctgccaag cacagcctgc tgtacgagta cttcaccgtg | 1560 |
| tacaacgagc tgaccaaagt gaaatacgtg accgagggaa tgcggaaagcc cgccttctg | 1620 |
| agcggcgagc agaaaaaaggc catcggttca ctgctgttca agaccaaccg gaaagtgacc | 1680 |
| gtgaagcagc tgaaaagagga ctacttcaag aaaatcgagt gttcgacag cgtggaaatc | 1740 |
| agcggcgtgg aagatcggtt caacgcctcc ctggcgccct atcacatct gctgaaaatt | 1800 |
| atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgaa agatatcg | 1860 |
| ctgaccctga cactgtttga ggaccggggc atgatcgagg aacggctgaa aacctatgcc | 1920 |
| cacctgtcg acgacaaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc | 1980 |
| aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg | 2040 |
| gatttcctga agtccgacgg cttcgccaac agaaaacttca tgcagctgat ccacgacgac | 2100 |
| agcctgacct ttaaagagga catccagaaa gcccaggtgt cggccaggg acactctctg | 2160 |
| cacgagcaga tcgccaatct ggccggatcc cccgcccatta agaagggcat cctgcagaca | 2220 |
| gtgaagattg tggacgagct cgtgaaagtg atggccaca agcccagaa catcgatc | 2280 |
| gaaatggcca gagagaacca gaccacccag aaggacaga agaacagccg cgagagaatg | 2340 |
| aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acacccctg | 2400 |
| gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctgcagaa tggcgggat | 2460 |
| atgtacgtgg accaggaact ggacatcaac cggctgtccg actacatgt ggaccacatt | 2520 |
| gtgccccagt cttcatcaa ggacgactcc atcgataaca aagtgtgac tcggagcgac | 2580 |
| aagaaccggg gcaagagcga caacgtccc tccgaagagg tcgtgaaagaa gatgaagaac | 2640 |
| tactggcgcc agctgctgaa tgccaaagctg attaccaga ggaagttcga caatctgacc | 2700 |
| aaggccgaga gaggcggcct gagcgaactg gataaggccg gttcattaa gcggcagctg | 2760 |
| gtggaaaccc ggcagatcac aaagcacgtg gcacagatcc tggactcccg gatgaacact | 2820 |
| aagtacgacg agaacgacaa actgatccgg gaagtgaaag tgcgttccct gaagtccaaag | 2880 |
| ctgggtgtccg acttcagaaa ggatttccag ttttacaaag tgcgcgagat caacaactac | 2940 |

| | |
|--------------------------------------------------------------------|------|
| caccacgccc acgacgccta cctgaacgcc gtcgtggaa ccgcctgat caaaaagtac | 3000 |
| cctaagctgg aaagcgagtt cgtgtacggc gattacaagg tgtacgacgt gcggaagatg | 3060 |
| atcgccaaga gcgagcagga aatcgcaag gctaccgcca agtacttctt ctacagcaac | 3120 |
| atcatgaact tttcaagac cgagatcaca ctggccaacg gcgagatcag aaagcggcct | 3180 |
| ctgatcgaga caaacggcga aaccggggag atcgtgtggg ataaggccg ggattttgcc | 3240 |
| acagtgcgga aagtgcgtc catgccccaa gtgaatatcg tgaaaaagac cgaggtgcag | 3300 |
| accggccgct tcagcaaaga gtctatcctg cccaagagga actccgacaa gctgatcgcc | 3360 |
| agaaagaagg attgggaccc taagaagtac ggcggcttg acagccccac cgtggcctac | 3420 |
| tctgtgctgg tggtgccaa agtggaaaag ggcaagtcca agaaactgaa gagtgtgaaa | 3480 |
| gagctgcgtgg ggatcaccat catggaaaga agcagcttcg agaagaatcc catcgacttt | 3540 |
| ctggaagcca agggctacaa agaagtgaaa aaggacctga tcatcaagct gcctaagtac | 3600 |
| tccctgttcg agctggaaaa cggccggaag cggatgcgtt cttctgcccgg cgaactgcag | 3660 |
| aagggaaacg agctggccct gccctccaaa tatgtgaact tccatgtacct ggccagccac | 3720 |
| tatgagaagc tgaagggctc ccccgaggat aatgagcaga aacagctgtt tgtggAACAG | 3780 |
| cacaagcaact acctggacga gatcatcgag cagattagcg agttctccaa ggcgtgtatc | 3840 |
| ctggccgatg ccaacctgga caaggtgcgtg agcgcctaca acaagcacgg ggataagccc | 3900 |
| atcagagagc aggccgagaa tatcatccac ctgtttaccc tgaccaacct gggagccct | 3960 |
| gccgccttca agtactttga caccaccatc gaccggaaaga ggtacaccag cacaaagag | 4020 |
| gtgctggacg ccaccctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac | 4080 |
| ctgtctcagc tgggaggcga ccccaagaaa aagcgcaaag tg | 4122 |

<210> 11

<211> 4764

<212> ADN

<213> Homo sapiens

5

<400> 11

| | |
|----------------------------------------------------------------------|-----|
| gcggggcggggc ggtgcgtatgt ccggagagga tggcccgccg gctggcccg gggcgccggc | 60 |
| ggcggatgcc cgggagcggc gacgggagca gctgcggcag tggggggcgc gggcgccgc | 120 |
| cgagcctggc cccggagagc gccgcgccc caccgtccgc ttgcagcgcg cgcgcgagtt | 180 |
| cctggcggcc tgtgcgggcg ggcacccgttgcgaggcgcgt ctgatgcgtc ggcgcgcccga | 240 |
| ccctggccccc ggcgcggagc tcgacccgcg cgcgcgcgcg cccgcgcgcg ccgtgcgttgc | 300 |
| ctccaccaac gccgacggta tcagcgcctt gcaccaggcgc agcgcgcgc gcccggcgtc | 360 |
| tcccggggcc aggtccaccc tctgcgtgcgc cacctggggc atcctcccttcc cccgttgcca | 420 |
| gtctcgatcc gccccgtcgt tccatggccctt gggctttgcc accctatgcgacacccgt | 480 |

| | | |
|-------|---------------------------------------------------------------|------|
| ccca | gtcccc cttaccattc ccattcgacc accccacttc cgaattggag ccgcttcaac | 540 |
| tggcc | ccttagccact ctgtgctgac cactctgcc caggcctact taccattccc | 600 |
| cttc | gaccta ctctcttccg cattggagtc gctttaactg gccctggatt tgccagcctg | 660 |
| tgctg | acccca tgca | 720 |
| gccc | cttc | 780 |
| gc | ccatccag ccggtcctgg actttgtctc ctccactgac ctgccc | 840 |
| cc | ctc | 900 |
| gg | cttc | 960 |
| cc | actgttt cccatccca ggcaggtaat gcttctatc acctgoattc tctccc | 1020 |
| gc | ctgtgccc cttctgtct gcagttgtg gcctgggtca cctctacggc tggccc | 1080 |
| cc | ttc | 1140 |
| gt | gttgc tccatccca aggtccgtc ttcc | 1200 |
| tg | tttccact cccttccccc ctgtgtgc tgcccaagga tgctttcc ggagca | 1260 |
| tg | atgtc ctc tgagcggatc ctccccgtgt ctgggtcctc tccgggc | 1320 |
| ca | cccaaccc catgcccgtc taactcgctg gttccatcc tcttataact tctgggg | 1380 |
| gt | ccatctc tcg | 1440 |
| cc | cccttctt gtaggcctgc atcatcaccc tt | 1500 |
| tc | ccctggctt tagccaccc tccatccatc tg | 1560 |
| tg | tggttgc ggtcac | 1620 |
| tc | atgtctg tccagcccc tgc | 1680 |
| tg | ccatccagg gatcctgtgt ccccgagctg ggaccac | 1740 |
| tg | ggctctg gttctggta ct | 1800 |
| ga | ccatccagg gacaggattg gtgacagaaa agccccatcc ttaggc | 1860 |
| tt | gggtctaa cccccaccc tcgtttagca gattc | 1920 |
| c | cttgc | 1980 |
| cc | ggccatccatc tgc | 2040 |
| cc | ggatcctggg agggagagct tggcagggggg tgggaggaa gggggggatg cgtgac | 2100 |
| cc | ccgtatggtg cg | 2160 |
| ag | cgatgttgg aaaaacaaaa tcagaataag ttgg | 2220 |
| c | cttcttagt ccccaattt tattgttcc | 2280 |
| cc | ccatggcca gccccgtcct ggcagggtcg tgg | 2340 |
| ct | ccctttgt gagaatggtg cgtc | 2400 |

ES 2 757 808 T3

| | |
|-----------------------------------------------------------------------|------|
| tttcttttc tccatccttc ttcccttaaa gagtccccag tgctatctgg gacatattcc | 2460 |
| tccgcccaga gcagggtccc gctccctaa ggccctgctc tgggcttctg ggtttgagtc | 2520 |
| cttggcaagc ccaggagagg cgctcaggct tccctgtccc ctttcctcggtt ccaccatctc | 2580 |
| atgcccctgg ctctcctgcc ccttcctac aggggttcct ggctctgctc ttcatcactga | 2640 |
| gccccgttcc cctgcattccc cgttcccttg catccccctt cccatgcata ccccaagggc | 2700 |
| cccaggccac ctacttggcc tggaccccac gagaggccac cccagccctg tctaccaggc | 2760 |
| tgcctttgg gtggatttctc ctccaactgt ggggtgactg cttggcaaac tcactcttgc | 2820 |
| gggtatccca ggaggcctgg agcattgggg tgggctgggg ttcagagagg agggattccc | 2880 |
| ttctcaggtt acgtggccaa gaagcaggggg agctgggtt gggtcaggc tgggtgtggg | 2940 |
| gtgaccagct tatgctgttt gccaggaca gccttagttt agcaactgaaa ccctcagtcc | 3000 |
| tagaaaaaca gggatggttt gtcactgtct ctgggtgact cttgattccc ggccagtttc | 3060 |
| tccacctggg gctgtgtttc tcgtcctgca tccttccttca ggcaggccc caagcatcgc | 3120 |
| ccccctgctg tggctgttcc caagttcttta gggtaaaaaa cgtgggttta tcaaccactt | 3180 |
| ggtgaggctg gtaccctgccc cccattcctg caccccaatt gccttagtgg cttaggggtt | 3240 |
| gggggctaga gtaggaggggg ctggagccag gattcttagg gctgaacaga gaagagctgg | 3300 |
| gggcctgggc tcctgggttt gagagaggag gggctggggc ctggactcct gggccggagg | 3360 |
| gaggagggggc tggggcctgg actcctgggt ctgagggtgg agggactggg ggcctggact | 3420 |
| cctgggtccg agggaggagg ggctggggcc tggactcgtg ggtctgaggg aggaggggt | 3480 |
| gggggctggg acttctgggt cttagggagg cggggctggg cctggacccc tgggtctgaa | 3540 |
| tggggagagg ctgggggcct ggactccttc atctgagggc ggaagggtcg gggccctggcc | 3600 |
| tcctgggttg aatggggagg ggttgggcct ggactctgga gtccctggc cccaggcctc | 3660 |
| aggcatctt cacagggatg cctgtactgg gcaggtcctt gaaaggaaaa ggcccatggc | 3720 |
| tctcctgccc cccctccccatcgc cactgggtt gaaataaacg agccgagttc | 3780 |
| atcccgttcc cagggcacgt gggcccccattt cacagcccgaa gtttccatgaa cctcatgctc | 3840 |
| ttggccctcg tagctccctc cggccctctc cagatggca gctttggaga ggtgagggac | 3900 |
| ttggggggta atttatccccg tggatctagg agtttagctt cactccttcc tcaagctccag | 3960 |
| ttcaggtccc ggagcccacc cagtgtccac aaggcctggg gcaagtcctt cctccgaccc | 4020 |
| cctggacttc ggctttgtc ccccaagtt ttggacccctt aaggaaagaa tgagaaacgg | 4080 |
| tggccctgtt cagccccctgg ctgcaggggcc ccgtgcagag gggccctcg tgaactggag | 4140 |
| tgtgacagcc tggggcccag gcacacaggt gtgcagctgt ctcacccctc tgggagtcoc | 4200 |
| gcccaggcccccttgc cttggccatggc cttccaccccttgc catagccctg | 4260 |

| | | |
|----|---------------------------------------------------------------------------------------------------|------|
| | ggcccacggc ttcgttcctg cagagtatct gctgggtgg tttccgagct tgacccttgg | 4320 |
| | aaggacctgg ctgggtttaa ggcaggaggg gctgggggcc aggactcctg gctctgaagg | 4380 |
| | aggagggct ggaacctctt ccctagtctg agcaactggaa gcgcacccctg tgggtggtga | 4440 |
| | cgggggtttt gccgtgtcta acaggtacca tgtgggttc ccgcacccag atgagaagcc | 4500 |
| | ccctcccttc cccgttcact tcctgtttgc agatagccag gagtccttc gtggttcca | 4560 |
| | ctgagactg aaggcctggc cggcctgacc actggcaac caggcgtatc ttAAACAGCC | 4620 |
| | agtggccaga ggctgttggg tcattttccc cactgtccta gcaccgtgtc cctggatctg | 4680 |
| | ttttcggtgc tccctcttgg a g t c c c g a c t t g t g g a c a c c g t g g t g g g g t a g g t g c g g | 4740 |
| | ctgacggctg tttccccccca ccag | 4764 |
| | <210> 12 | |
| | <211> 42 | |
| 5 | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 12 | |
| | accccacagu gggccacua guuuuagagc uaugcuguuu ug | 42 |
| 10 | <210> 13 | |
| | <211> 86 | |
| | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| 15 | <223> SINTETIZADO | |
| | <400> 13 | |
| | ggaaccaauuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa | 60 |
| | aguggcacccg agucggugcu uuuuuu | 86 |
| | <210> 14 | |
| | <211> 62 | |
| 20 | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 14 | |
| | accccacagu gggccacua guuuuagagc uagaaaauagc aaguuaaaaau aaggcuaguc | 60 |
| 25 | cg | 62 |
| | <210> 15 | |
| | <211> 25 | |
| | <212> ADN | |
| | <213> Secuencia artificial | |
| 30 | <220> | |
| | <223> SINTETIZADO | |
| | <400> 15 | |
| | ctgacctt ctcttcctcc cacag | 25 |

ES 2 757 808 T3

| | | |
|---------------------------------------------------------------------|--|------|
| <210> 16 | | |
| <211> 1009 | | |
| <212> ADN | | |
| <213> Secuencia artificial | | |
| 5 <220> | | |
| <223> SINTETIZADO | | |
| <400> 16 | | |
| gccaccatgg actacaaaga cgatgacgac aaggtcgact ctagagctgc agagagcgcac | | 60 |
| gagagcggcc tgcccgccat ggagatcgag tgccgcata ccggcaccct gaacggcgtg | | 120 |
| gagttcgagc tggtgggcgg cggagagggc accccccgagc agggccgcat gaccaacaag | | 180 |
| atgaagagca ccaaaggcgc cctgacccccc agccccctacc tgctgagcca cgtgatggc | | 240 |
| tacggcttctt accacttcgg cacctacccc agcggctacg agaaccctt cctgcacgccc | | 300 |
| atcaacaacg gcccgtacac caacacccgc atcgagaagt acgaggacgg cggcgtgctg | | 360 |
| cacgtgagct tcagctacccg ctacgaggcc ggccgcgtga tcggcactt caaggtgatg | | 420 |
| ggcaccggct tccccgagga cagcgtgatc ttccaccgaca agatcgtccg cagcaacgccc | | 480 |
| accgtggagc acctgcaccc catggcgtat aacgatctgg atggcagctt caccgcacc | | 540 |
| ttcagcctgc gcgacggcgg ctactacagc tccgtggtgg acagccacat gcacttcaag | | 600 |
| agcgccatcc accccagcat cctgcagaac gggggccccca tgttcgcctt ccgcgcgtg | | 660 |
| gaggaggatc acagcaacac cgagctggc atcgtggagt accagcacgc cttcaagacc | | 720 |
| ccggatgcag atgccggtga agaatgaaga tctctgtgcc ttctagttgc cagccatctg | | 780 |
| ttgtttgccc ctccccgtg cttcccttga ccctggaagg tgccactccc actgtcctt | | 840 |
| cctaataaaa tgagggaaatt gcatcgatt gtctgagtag gtgtcattct attctggggg | | 900 |
| gtgggggtgg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg | | 960 |
| atgcgggtgg ctctatggac tcgagggtta aacgtcgacg cggccgcgt | | 1009 |
| 10 <210> 17 | | |
| <211> 355 | | |
| <212> PRT | | |
| <213> Secuencia artificial | | |
| <220> | | |
| <223> SINTETIZADO | | |
| 15 <400> 17 | | |

Met Ser Gly Glu Asp Gly Pro Ala Ala Gly Pro Gly Ala Ala Ala Ala

ES 2 757 808 T3

| | | | |
|-----------------------------------------------------------------|-----|-----|-----|
| 1 | 5 | 10 | 15 |
| Ala Ala Arg Glu Arg Arg Arg Glu Gln Leu Arg Gln Trp Gly Ala Arg | | | |
| 20 | 25 | 30 | |
| Ala Gly Ala Glu Pro Gly Pro Gly Glu Arg Arg Ala Arg Thr Val Arg | | | |
| 35 | 40 | 45 | |
| Phe Glu Arg Ala Ala Glu Phe Leu Ala Ala Cys Ala Gly Gly Asp Leu | | | |
| 50 | 55 | 60 | |
| Asp Glu Ala Arg Leu Met Leu Arg Ala Ala Asp Pro Gly Pro Gly Ala | | | |
| 65 | 70 | 75 | 80 |
| Glu Leu Asp Pro Ala Ala Pro Pro Pro Ala Arg Ala Val Leu Asp Ser | | | |
| 85 | 90 | 95 | |
| Thr Asn Ala Asp Gly Ile Ser Ala Leu His Gln Ala Thr Met Asp Tyr | | | |
| 100 | 105 | 110 | |
| Lys Asp Asp Asp Asp Lys Val Asp Ser Arg Ala Ala Glu Ser Asp Glu | | | |
| 115 | 120 | 125 | |
| Ser Gly Leu Pro Ala Met Glu Ile Glu Cys Arg Ile Thr Gly Thr Leu | | | |
| 130 | 135 | 140 | |
| Asn Gly Val Glu Phe Glu Leu Val Gly Gly Glu Gly Thr Pro Glu | | | |
| 145 | 150 | 155 | 160 |
| Gln Gly Arg Met Thr Asn Lys Met Lys Ser Thr Lys Gly Ala Leu Thr | | | |
| 165 | 170 | 175 | |
| Phe Ser Pro Tyr Leu Leu Ser His Val Met Gly Tyr Gly Phe Tyr His | | | |
| 180 | 185 | 190 | |
| Phe Gly Thr Tyr Pro Ser Gly Tyr Glu Asn Pro Phe Leu His Ala Ile | | | |
| 195 | 200 | 205 | |
| Asn Asn Gly Gly Tyr Thr Asn Thr Arg Ile Glu Lys Tyr Glu Asp Gly | | | |
| 210 | 215 | 220 | |
| Gly Val Leu His Val Ser Phe Ser Tyr Arg Tyr Glu Ala Gly Arg Val | | | |
| 225 | 230 | 235 | 240 |
| Ile Gly Asp Phe Lys Val Met Gly Thr Gly Phe Pro Glu Asp Ser Val | | | |
| 245 | 250 | 255 | |

ES 2 757 808 T3

Ile Phe Thr Asp Lys Ile Val Arg Ser Asn Ala Thr Val Glu His Leu
260 265 270

His Pro Met Gly Asp Asn Asp Leu Asp Gly Ser Phe Thr Arg Thr Phe
275 280 285

Ser Leu Arg Asp Gly Gly Tyr Tyr Ser Ser Val Val Asp Ser His Met
290 295 300

His Phe Lys Ser Ala Ile His Pro Ser Ile Leu Gln Asn Gly Gly Pro
305 310 315 320

Met Phe Ala Phe Arg Arg Val Glu Glu Asp His Ser Asn Thr Glu Leu
325 330 335

Gly Ile Val Glu Tyr Gln His Ala Phe Lys Thr Pro Asp Ala Asp Ala
340 345 350

Gly Glu Glu
355

<210> 18

<211> 21

<212> ADN

5 <213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 18

ccactctgtc ctgaccacat t

21

10 <210> 19

<211> 17

<212> ADN

<213> Secuencia artificial

<220>

<223> SINTETIZADO

<400> 19

gcggcactcg atctcca

17

<210> 20

<211> 711

20 <212> ADN

<213> Mus musculus

<400> 20

| | | |
|----|---------------------------------------------------------------------|-----|
| | gagcggctgc gggcggtg caagcacgtt tccgacttga gttgcctcaa gaggggcgtg | 60 |
| | ctgagccaga cctccatcgc gcactccggg gagtggaggg aaggagcgag ggctcagttg | 120 |
| | ggctgtttg gaggcagaa gcacttgctc tcccaaagtc gctctgagtt gttatcagta | 180 |
| | agggagctgc agtggagtag gcggggagaa ggccgcaccc ttatccggag gggggagggg | 240 |
| | agtgttgc aa taccttctg ggagttctct gctgcctcct ggcttctgag gaccgcctg | 300 |
| | ggcctggag aatcccttcc ccctcttccc tcgtgatctg caactccagt ctttctagaa | 360 |
| | gatggcggg agtcttctgg gcaggctaa aggctaacct ggtgtgtggg cgttgtcctg | 420 |
| | caggggaatt gaacagggtgt aaaattggag ggacaagact tcccacagat ttcgggtttt | 480 |
| | gtcgggaagt ttttaata tag gggcaaataa gaaaaatggg aggataggta gtcatctggg | 540 |
| | gttttatgca gcaaaactac aggttattat tgcttgtat ccgcctcgga gtatttcca | 600 |
| | tcgaggtaga ttaaagacat gtcacccga gtttatact ctctgcttg agatccttac | 660 |
| | tacagtatga aattacagtg tcgcgagtt gactatgtaa gcagaatttt a | 711 |
| | <210> 21 | |
| | <211> 42 | |
| | <212> ARN | |
| 5 | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 21 | |
| | cuccagucuu ucuagaagau guuuuagagc uaugcuguuu ug | 42 |
| 10 | <210> 22 | |
| | <211> 42 | |
| | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| 15 | <223> SINTETIZADO | |
| | <400> 22 | |
| | ugaacaggug uaaaauugga guuuuagagc uaugcuguuu ug | 42 |
| | <210> 23 | |
| | <211> 42 | |
| 20 | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 23 | |
| 25 | ugucgggaag uuuuuuaaua guuuuagagc uaugcuguuu ug | 42 |
| | <210> 24 | |
| | <211> 642 | |
| | <212> ADN | |
| | <213> Rattus rattus | |
| 30 | <400> 24 | |

| | | |
|----|-------------------------------------------------------------------|-----|
| | gggattcctc cttgagttgt ggcactgagg aacgtgctga acaagaccta cattgcactc | 60 |
| | cagggagtgg atgaaggagt tggggctcag tcgggttgc ttggagacaa gaagcacttg | 120 |
| | ctctccaaaa gtcggttga gttatcatta agggagctgc agtggagtag gcggagaaaa | 180 |
| | ggccgcaccc ttctcaggac gggggagggg agtgttgc aa taccttctg ggagttctct | 240 |
| | gctgcctcct gtctctgag gaccgcctg ggcctggaag attcccttcc cccttcttcc | 300 |
| | ctcggtatct gcaactggag tctttctgga agataggcgg gagtcttctg ggcaggctta | 360 |
| | aaggctaacc tggtgcgtgg ggcgttgc tgcagaggaa ttgaacaggt gtaaaaattgg | 420 |
| | agggccaaga ctccccacag atttcgatt gtgttgttaa gtatttaat aggggcaaata | 480 |
| | aaggaaata gactaggcac tcacctgggg ttttatgcag caaaactaca ggttattttt | 540 |
| | gctgtgatc cgccctggag aattttcac cgaggttagat tgaagacatg cccacccaaa | 600 |
| | tttaatatt cttccacttg cgatccttgc tacagtatga aa | 642 |
| | <210> 25 | |
| | <211> 42 | |
| | <212> ARN | |
| 5 | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 25 | |
| | aggggaagg gaaucuuucca guuuuagagc uaugcuguuu ug | 42 |
| 10 | <210> 26 | |
| | <211> 42 | |
| | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| 15 | <223> SINTETIZADO | |
| | <400> 26 | |
| | ucugcaacug gagucuuucu guuuuagagc uaugcuguuu ug | 42 |
| | <210> 27 | |
| | <211> 42 | |
| 20 | <212> ARN | |
| | <213> Secuencia artificial | |
| | <220> | |
| | <223> SINTETIZADO | |
| | <400> 27 | |
| 25 | aggcgggagu cuucugggca guuuuagagc uaugcuguuu ug | 42 |

REIVINDICACIONES

1. Un complejo de endonucleasa guiado por ARN diseñado por ingeniería genética que comprende:

un ARN guía que comprende

- 5 (i) una primera región complementaria a un sitio diana en una secuencia cromosómica eucariota que puede emparejarse formando pares de bases con el sitio diana, que comprende de aproximadamente 10 nucleótidos a más de aproximadamente 25 nucleótidos,
 - (ii) una segunda región que forma una estructura de tallo y bucle, y
 - (iii) una tercera región que es fundamentalmente monocatenaria,
- 10 en el que (i), (ii) y (iii) están dispuestos en la dirección 5' a 3', y el ARN guía comprende dos moléculas separadas,

en el que se forma un complejo proteína-ARN entre el ARN guía y una proteína CRISPR/Cas9 de tipo II, que comprende además una señal de localización nuclear, y el ARN guía interactúa con la proteína CRISPR/Cas9 de tipo II para guiar a la proteína al sitio diana específico.

15 2. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 1, en el que la primera molécula del ARN guía comprende la primera región del ARN guía y la mitad del tallo de la segunda región del ARN guía.

3. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 1 o la reivindicación 2, en el que la segunda molécula del ARN guía comprende la otra mitad del tallo de la segunda región del ARN guía y la tercera región del ARN guía.

20 4. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquiera de las reivindicaciones 1-3, en el que la segunda región del ARN guía tiene una longitud de aproximadamente 16 a aproximadamente 60 nucleótidos.

25 5. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquiera de las reivindicaciones 1-4, en el que la tercera región del ARN guía tiene una longitud de aproximadamente 5 a aproximadamente 60 nucleótidos.

6. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquiera de las reivindicaciones 1-5, en el que la proteína CRISPR/Cas9 de tipo II comprende solo un dominio nucleasa funcional.

7. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 6, en el que la proteína CRISPR/Cas9 de tipo II comprende un dominio tipo RuvC no funcional.

30 8. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 6, en el que la proteína CRISPR/Cas9 de tipo II comprende un dominio tipo HNH no funcional.

9. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior, en el que la proteína CRISPR/Cas9 tipo II es de una especie de *Streptococcus*.

35 10. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 9, en el que la proteína CRISPR/Cas9 tipo II es de *Streptococcus pyogenes*.

11. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior, en el que la NLS está localizada en el extremo C de la endonucleasa guiada por ARN.

12. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior, en el que la proteína CRISPR/Cas9 de tipo II está codificada por ADN optimizado para la expresión en eucariotas.

40 13. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 12, en el que el ADN está unido de forma operativa a una secuencia promotora control para la expresión en eucariotas.

14. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 12 o la reivindicación 13, en el que el ADN que codifica la endonucleasa guiada por ARN diseñada por ingeniería genética está comprendido en un vector.

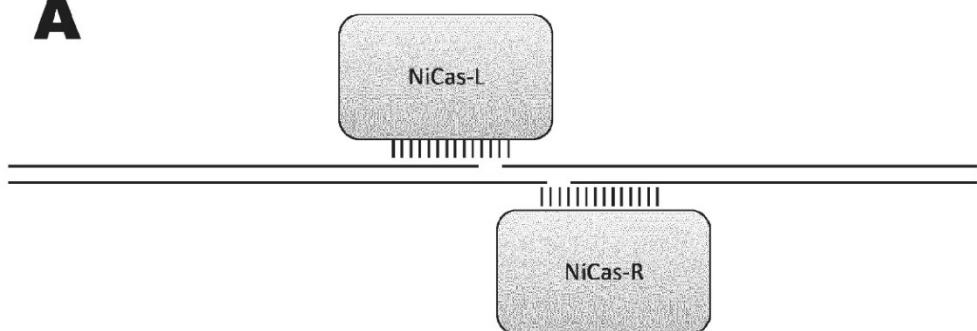
45 15. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior, en el que el ARN guía está codificado por ADN unido de forma operativa a una secuencia promotora control para la expresión en eucariotas.

16. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 15, en el que el ADN que codifica el ARN guía está comprendido en un vector.

50 17. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior,

- en el que el sitio diana está corriente arriba de un motivo adyacente protoespaciador (PAM).
18. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 17, en el que el PAM está inmediatamente corriente abajo del sitio diana.
19. El complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de la reivindicación 17 o la 5 reivindicación 18, en el que el PAM es NGG o NGNG, en el que N se define como cualquier nucleótido.
20. Uso del complejo de endonucleasa guiado por ARN diseñado por ingeniería genética de cualquier reivindicación anterior, y opcionalmente al menos un polinucleótido donador que comprende una secuencia donadora, para modificar 10 una secuencia cromosómica, en el que el uso no comprende un procedimiento para modificar la identidad genética de la línea germinal de un ser humano, en el que el uso no comprende el uso de embriones humanos para fines industriales o comerciales y, en el que dicho procedimiento no comprende un procedimiento para el tratamiento del cuerpo humano o animal mediante cirugía 15 o terapia.
21. El uso de la reivindicación 20, en el que la secuencia cromosómica se repara mediante un proceso de reparación de ADN, de tal manera que la secuencia cromosómica se modifica mediante la delección de al menos un nucleótido, una inserción de al menos un nucleótido, una sustitución de al menos un nucleótido, o una combinación de las mismas.
22. El uso de acuerdo con la reivindicación 21, en el que el proceso de reparación de ADN es un proceso de reparación 20 de unión de extremos no homólogos (NHEJ).
23. El uso de acuerdo con la reivindicación 21, que comprende además un polinucleótido donador que comprende 25 secuencias corriente arriba y corriente abajo, en el que el proceso de reparación del ADN es una reparación dirigida por homología (HDR) y la secuencia donadora se integra en la secuencia cromosómica.
24. El uso de cualquiera de las reivindicaciones 20-23, en el que la secuencia cromosómica está en una célula eucariota, en el que la célula eucariota es una célula humana, una célula de mamífero no humano o una célula de vertebrado no mamífero.
25. El uso de cualquiera de las reivindicaciones 20-23, en el que la secuencia cromosómica está en una célula 25 eucariota, en el que la célula eucariota es una célula de invertebrado, una célula de insecto, una célula vegetal, una célula de levadura o un organismo eucariota unicelular.
26. El uso de la reivindicación 25, en el que la célula eucariota es una célula vegetal.
27. El uso de cualquiera de las reivindicaciones 20-23 en el que la secuencia cromosómica está en un embrión de mamífero no humano.

A



B

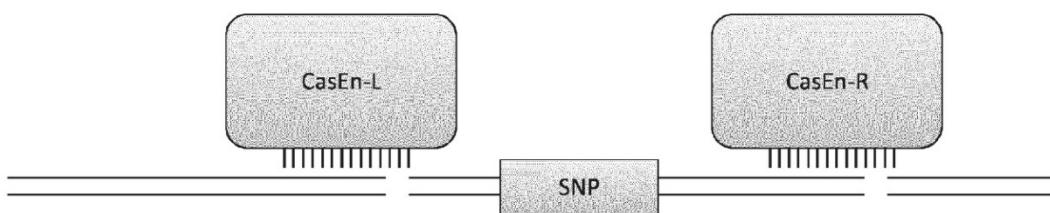


FIG. 1

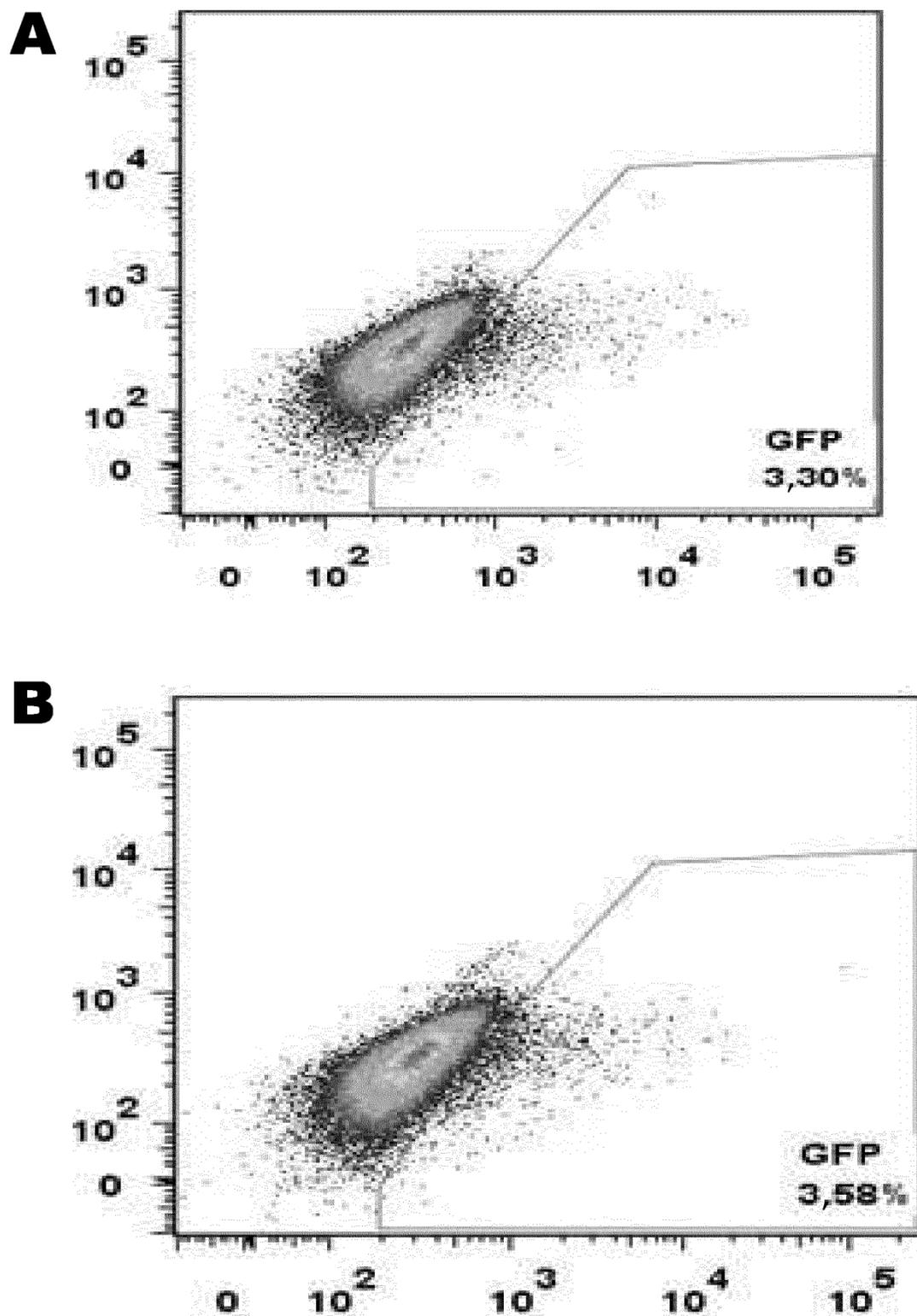


FIG. 2

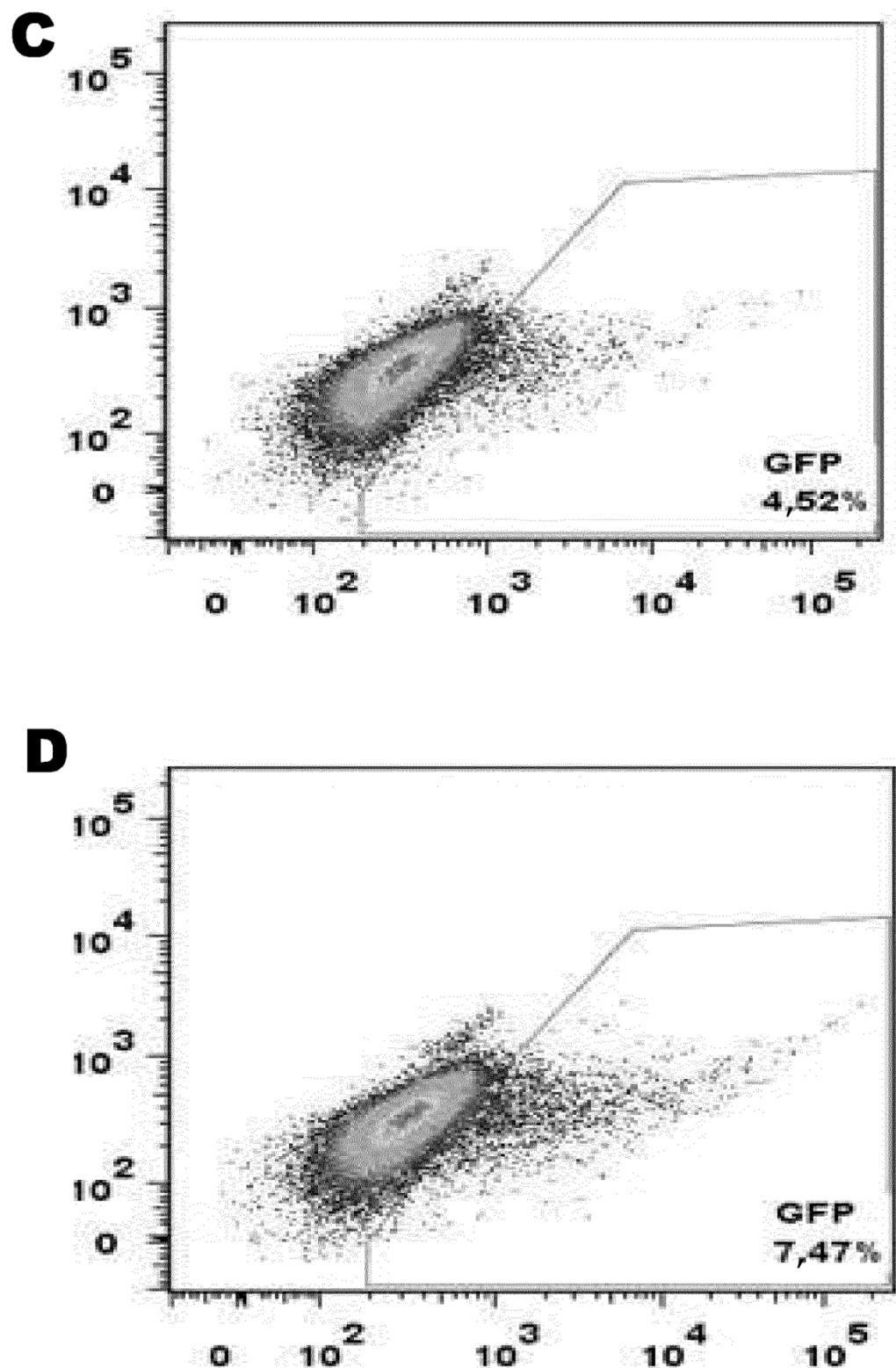


FIG. 2

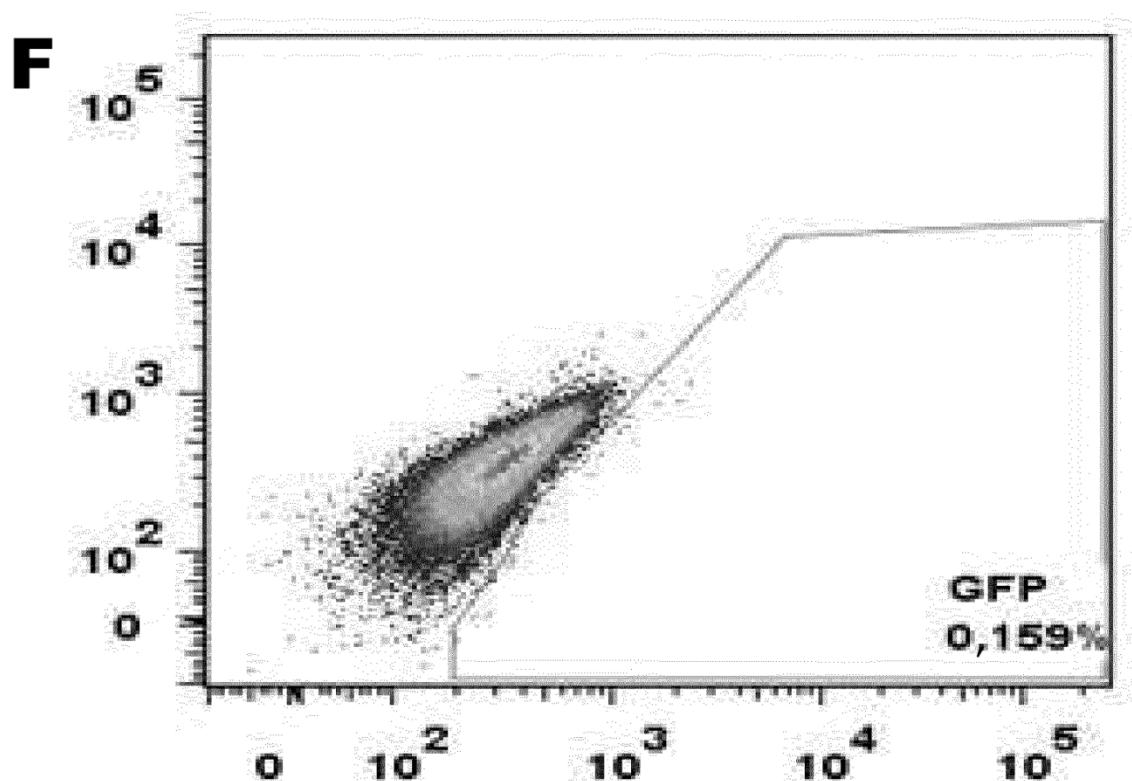
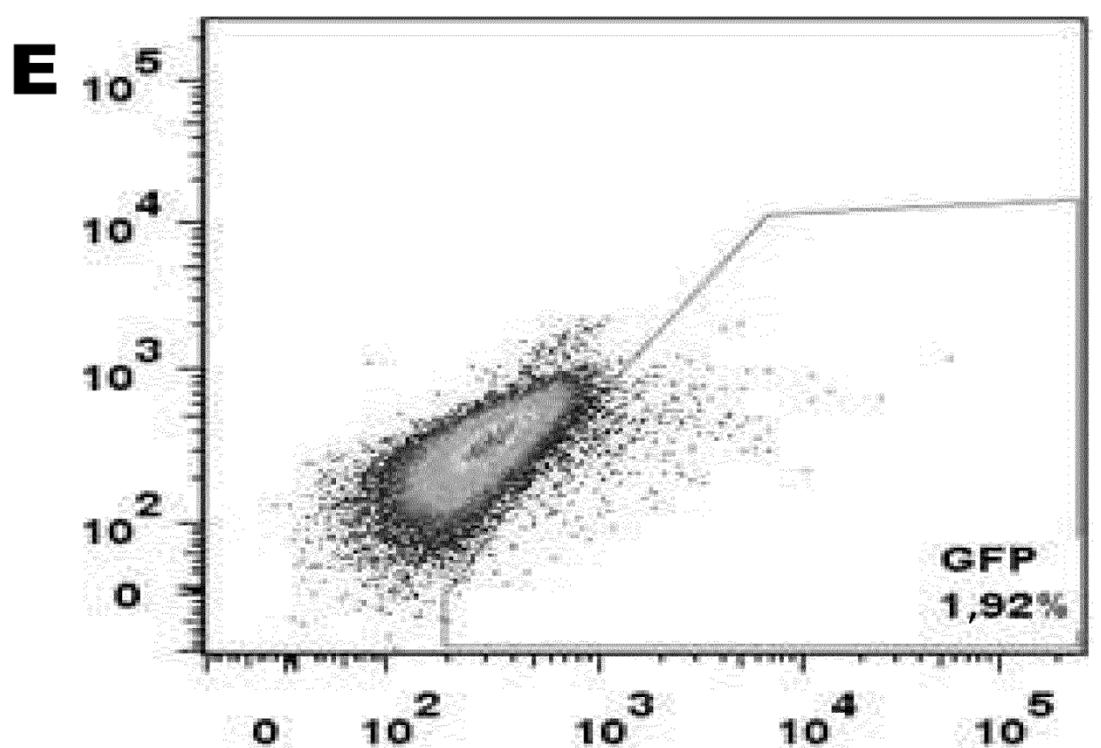


FIG. 2

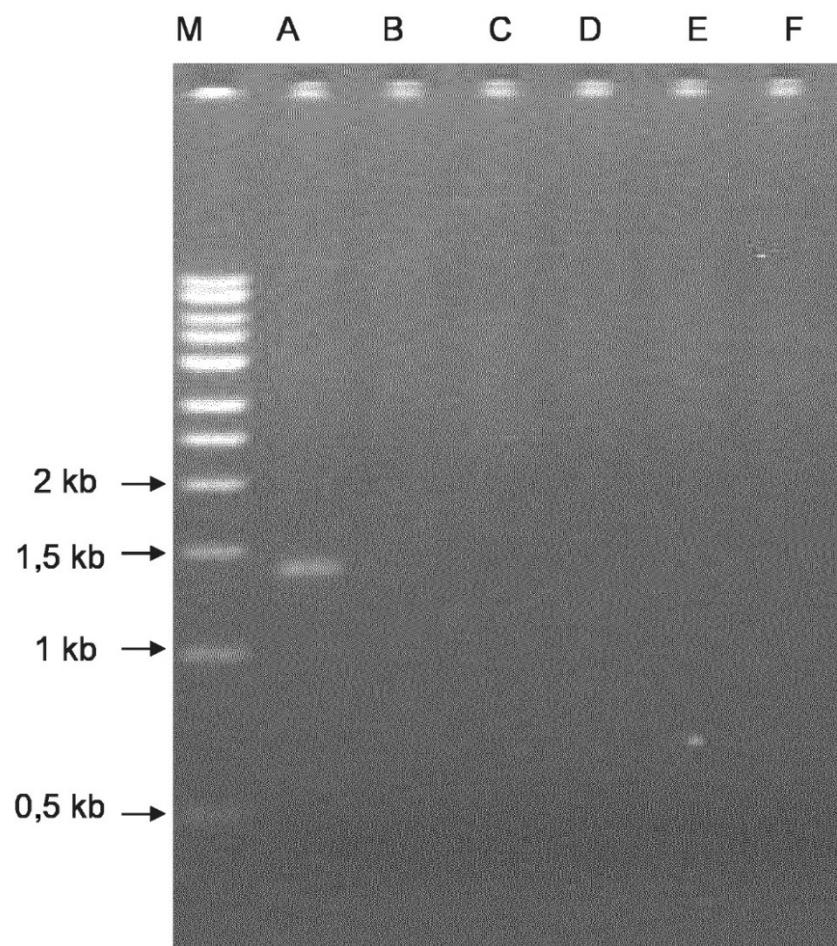


FIG. 3