

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 713 243**

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 16183725 (7)**

97 Fecha y número de publicación de la concesión europea: **05.12.2018 EP 3138912**

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:

20.05.2019

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 2 713 243 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 una modificación dirigida del genoma. En particular, la divulgación se refiere a endonucleasas guiadas por ARN que comprenden la proteína de tipo CRISPR/Cas y procedimientos para usar dichas proteínas para modificar y regular secuencias cromosómicas dirigidas.

Antecedentes de la invención

10 La modificación genómica dirigida es una herramienta poderosa para la manipulación genética de las células eucariotas, de embriones y de animales. Por ejemplo, las secuencias exógenas se pueden integrar en localizaciones genómicas dirigidas y/o las secuencias cromosómicas endógenas específicas se pueden eliminar, desactivar o modificar. Los procedimientos actuales se basan en el uso de enzimas nucleasas diseñadas genéticamente, tales como, por ejemplo, las nucleasas de dedos de cinc (ZFN, del inglés *zinc finger nucleases*) o nucleasas de tipo activadores de transcripción (TALEN, del inglés *transcription activator-like effector nucleases*). Estas nucleasas quiméricas contienen módulos de unión a ADN programables y específicos de secuencia unidos a un dominio de escisión de ADN no específico. Cada nueva diana genómica, sin embargo, requiere el diseño de una nueva ZFN o TALEN que comprende un nuevo módulo de unión a ADN específico de secuencia. Por lo tanto, estas nucleasas diseñadas a medida tienden a ser costosas y requieren mucho tiempo para prepararse. Por otra parte, las especificidades de ZFN y TALEN son tales que pueden mediar en escisiones fuera de la diana.

20 Por lo tanto, existe una necesidad de una tecnología de modificación de genoma dirigida que no requiera el diseño de una nueva nucleasa para cada nueva localización genómica dirigida. Además, existe una necesidad de una tecnología con especificidad aumentada con pocos o sin efectos fuera de la diana.

Sumario de la invención

25 La presente invención proporciona un procedimiento para modificar una secuencia cromosómica en una célula eucariota, comprendiendo el procedimiento introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas y, opcionalmente un polinucleótido donador en el que cada sistema de nickasa guiado por ARN comprende (i) una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble y (ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN, en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protospaciador (PAM), y los sitios diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y cultivar la célula eucariota, de manera que las dos endonucleasas guiadas por ARN escindan cadenas opuestas de la secuencia cromosómica en estrecha proximidad, lo suficiente como para introducir una rotura de doble cadena en la secuencia cromosómica, y reparar la rotura de cadena doble mediante un proceso de reparación de ADN que lleva a la modificación de la secuencia cromosómica, en la que el procedimiento no comprende un proceso de 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.

40 La presente invención también proporciona un procedimiento ex vivo o in vitro para modificar una secuencia cromosómica de una célula eucariota, comprendiendo el procedimiento:

a) introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas y, opcionalmente, un polinucleótido donador en el que cada sistema de nickasa guiado por ARN comprende

45 (i) Una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble; y
(ii) Un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN,

50 en la que cada sitio diana está seguida inmediatamente por un motivo adyacente de protospaciador (PAM) y los sitios diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y

55 b) Cultivar la célula eucariota de tal manera que las dos endonucleasas guiadas por ARN escindan cadenas opuestas de la secuencia cromosómica en estrecha proximidad, lo suficiente como para introducir una rotura de doble cadena en la secuencia cromosómica, y reparar la rotura de doble cadena mediante un proceso de reparación de ADN que lleva a la modificación de la secuencia cromosómica, en la que el procedimiento no comprende un proceso de modificar la identidad genética de la línea germinal de un ser humano.

En una realización, la endonucleasa guiada por ARN puede derivar de una proteína Cas9. En otra realización, el ácido nucleico que codifica la endonucleasa guiada por ARN introducida en la célula o embrión puede ser ARNm. En una realización adicional, el ácido nucleico que codifica la endonucleasa guiada por ARN introducida en la célula puede ser ADN. En una realización adicional, el ADN que codifica la endonucleasa guiada por ARN puede ser parte de un vector que comprende adicionalmente una secuencia que codifica el ARN guía. En determinadas realizaciones, la célula eucariota puede ser una célula humana, una célula de mamífero no humano, una célula madre, una célula de vertebrado no mamífero, una célula de invertebrado, una célula vegetal, o un organismo eucariota unicelular. En otras determinadas realizaciones, la célula eucariota es un embrión unicelular de animal no humano.

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

Breve descripción de los dibujos

La **FIG. 1** representa el gráfico de la modificación del genoma usando dos endonucleasas guiadas por ARN, en la que una rotura de doble cadena se genera mediante dos endonucleasas guiadas por ARN que se han convertido en nickasas.

La **FIG. 2** el clasificador de células activadas por fluorescencia (FACS) de células K562 humanas transfectadas con ácido nucleico de Cas9, ARN que guía Cas9 y donador de ADN AAVS1-GFP. El eje Y representa la intensidad de auto fluorescencia en un canal rojo, y el eje X representa la intensidad de fluorescencia verde. **(A)** células K562 transfectadas con 10 µg de ARNm de Cas9 transcrito con un análogo de casquete anti-inverso, 0,3 nmol de doble cadena de ARNcr-ARNtracr prealineada, y 10 µg ADN del plásmido AAVS1-GFP; **(B)** células K562 transfectadas con 10 µg de ARNm de Cas9 transcrito con un análogo de casquete anti-inverso, 0,3 nmol de ARN quimérico y 10 µg ADN del plásmido AAVS1-GFP; **(C)** células K562 transfectadas con 10 µg de ARNm de Cas9 con casquete mediante la reacción de postranscripción de casquete, 0,3 nmol de ARN quimérico y 10 µg ADN del plásmido AAVS1-GFP; **(D)** células K562 transfectadas con 10 µg de ADN del plásmido de Cas9, 5 µg de ADN del plásmido de ARN quimérico con U6, y 10 µg de ADN del plásmido AAVS1-GFP; **(E)** células K562 transfectadas con 10 µg de ADN del plásmido AAVS1-GFP; **(F)** células K562 transfectadas solo con reactivos de transfección.

La **FIG. 3** presenta un análisis de PCR de unión que documenta la integración dirigida de 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 transcrito con un análogo de casquete anti-inverso, 0,3 nmol de doble cadena de ARNcr-ARNtracr prealineada, y 10 µg ADN del plásmido AAVS1-GFP; Carril B: células K562 transfectadas con 10 µg de ARNm de Cas9 transcrito con un análogo de casquete anti-inverso, 0,3 nmol de ARN quimérico y 10 µg ADN del plásmido AAVS1-GFP; Carril C: células K562 transfectadas con 10 µg de ARNm de Cas9 con casquete mediante la reacción de postranscripción de casquete, 0,3 nmol de ARN quimérico y 10 µg ADN del plásmido AAVS1-GFP; Carril D: células K562 transfectadas con 10 µg de ADN del plásmido de Cas9, 5 µg de ADN del plásmido de ARN quimérico con U6, y 10 µg de ADN del plásmido AAVS1-GFP; Carril E: células K562 transfectadas con 10 µg de ADN del plásmido AAVS1-GFP; Carril F: células K562 transfectadas solo con reactivos de transfección.

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 interacciona con un ARN guía para dirigir la endonucleasa a una secuencia de nucleótidos específica para la escisión. También se desvelan ácidos nucleicos que codifican las endonucleasas guiadas por ARN, así como procedimientos para usar las endonucleasas guiadas por ARN para modificar las secuencias cromosómicas de células eucariotas o embriones. La endonucleasa guiada por ARN interacciona con los ARN guía específicos, cada uno de los cuales dirige la endonucleasa a un sitio específico dirigido, en cuyo sitio la endonucleasa guiada por ARN introduce una rotura de doble cadena que se puede reparar mediante un proceso de reparación de ADN de manera que la secuencia cromosómica se modifica. Dado que la especificidad la proporciona el ARN guía, la endonucleasa basada en ARN es universal y se puede usar con diferentes ARN guías para dirigirse a diferentes secuencias genómicas. Los procedimientos desvelados en el presente documento se pueden usar para dirigirse y modificar secuencias cromosómicas específicas y/o introducir secuencias exógenas en localizaciones diana en el genoma de células o embriones. Se excluyen los procedimientos que comprenden un proceso para modificar la identidad genética de la línea germinal de un ser humano. Por otro lado, el direccionamiento es específico con efectos limitados fuera de la diana.

(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 el núcleo de células eucariotas y embriones tales como, por ejemplo, embriones unicelulares no humanos. Las endonucleasas guiadas por ARN también comprenden al menos un dominio nucleasa y al menos un dominio que interacciona con un ARN guía. Una endonucleasa guiada por ARN se dirige a una secuencia específica de ácido nucleico (o sitio diana) mediante un ARN guía. El ARN guía interacciona 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 es capaz de introducir una rotura de doble

cadena en el sitio diana de la secuencia de ácido nucleico. Dado que el ARN guía proporciona la especificidad para la escisión dirigida, la endonucleasa de la endonucleasa guiada por ARN es universal y se puede usar con diferentes ARN guía para escindir diferentes secuencias diana de ácido nucleico. En el presente documento se desvelan endonucleasas guiadas por ARN, ácidos nucleicos aislados (es decir, ARN y ADN) que codifican las endonucleasas guiadas por ARN, los vectores que comprenden ácidos nucleicos que codifican las endonucleasas guiadas por ARN y los complejos de proteína-ARN que comprenden la endonucleasa guiada por ARN más un ARN guía.

La endonucleasa guiada por ARN puede provenir de un sistema de repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR)/(Cas) asociado a CRISPR. 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 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.

En una realización, la endonucleasa guiada por ARN proviene de un sistema CRISPR/Cas de tipo II. En realizaciones específicas, la endonucleasa guiada por ARN deriva de una proteína Cas9. La proteína Cas9 puede ser de *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Nocardiaopsis 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.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulfurudis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochrochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosipho 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 de unión a ARN interactúan con los ARN guía. Las proteínas CRISPR/Cas también pueden comprender dominios nucleasa (es decir, dominios DNasa o RNasa), dominios de unión a ADN, dominios helicasa, dominios RNasa, dominios de interacción proteína-proteína, dominios de dimerización, así como otros dominios.

La proteína de tipo CRISPR/Cas puede ser una proteína CRISPR/Cas de tipo silvestre, una proteína CRISPR/Cas modificada, o un fragmento de una proteína CRISPR/Cas de tipo silvestre o modificada. La proteína de tipo CRISPR/Cas se puede modificar para aumentar la afinidad y/o especificidad de unión a ácido nucleico, alterar una actividad enzimática y/o cambiar otra propiedad de la proteína. Por ejemplo, los dominios nucleasa (es decir, DNasa, RNasa) de la proteína de tipo CRISPR/Cas se pueden modificar, eliminar o desactivar. Como alternativa, se puede truncar la proteína de tipo CRISPR/Cas para retirar dominios que no son esenciales para la función de la proteína de fusión. La proteína de tipo CRISPR/Cas también se puede truncar o modificar para optimizar la actividad del dominio efector de la proteína de fusión.

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

En general, una proteína Cas9 comprende al menos dos dominios de nucleasa (es decir, DNasa). Por ejemplo, una proteína Cas9 puede comprender un dominio de nucleasa de tipo RuvC y un dominio de nucleasa de tipo HNH. Los dominios RuvC y HNH funcionan juntos para cortar cadenas simples para generar una rotura de cadena doble en el ADN (Jinek y col., Science, 337: 816-821). En la presente invención, la proteína derivada de Cas-9 se modifica para que contenga solo un dominio de nucleasa funcional (bien un dominio de nucleasa de tipo RuvC o de tipo HNH). Por ejemplo, la proteína derivada de Cas9 se puede modificar, de manera que uno de los dominios de nucleasa se deleciona o se muta, de manera que ya no sea funcional (es decir, la actividad nucleasa está ausente). En la presente invención, la proteína derivada de Cas9 es capaz de introducir una incisión en un ácido nucleico de cadena doble (tal proteína se denomina "nickasa" (del inglés *nick*, incisión)), pero no escinde el ADN de cadena doble. Por ejemplo, una conversión de aspartato a alanina (D10A) en un dominio de tipo Ruv convierte la proteína derivada de Cas9 en una nickasa. Del mismo modo, una conversión de histidina a alanina (H840A o H839A) en un dominio HNH convierte la proteína derivada de Cas9 en una nickasa. Cada dominio de nucleasa se puede modificar usando procedimientos bien conocidos, tal como la mutagénesis de sitio dirigido, la mutagénesis mediada por PCR y la síntesis génica total, así como otros procedimientos conocidos en la materia.

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

La endonucleasa guiada por ARN puede comprender adicionalmente al menos un dominio de penetración celular. El dominio de penetración celular puede ser una secuencia de péptido de penetración celular que proviene de la proteína TAT del VIH-1. A modo de ejemplo, la secuencia de penetración celular de TAT puede ser GRKKRRQRRPPQPKKKRKV (SEQ ID NO:4). Como alternativa, el dominio de penetración celular puede ser TLM (PLSSIFSRIGDPPKPKKKRKV; SEQ ID NO:5, una secuencia de péptido de penetración celular que proviene del virus de la hepatitis B. En otra realización alternativa, el dominio de penetración celular puede ser MPG (GALFLGWLGAAGSTMGAPKKKKRKV; SEQ ID NO:6 o GALFLGFLGAAGSTMGAWSQPKKKRKV; 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 de péptido de poliarginina. El dominio de penetración celular se puede localizar en el extremo N-terminal, en el extremo C-terminal o en una localización interna de la proteína.

La endonucleasa guiada por ARN también puede comprender adicionalmente al menos un dominio marcador. Los ejemplos no limitantes de dominios marcadores incluyen proteínas fluorescentes, etiquetas de purificación y etiquetas de epítipo. En algunas realizaciones, el dominio marcador puede ser una proteína fluorescente. Los ejemplos no limitantes de proteínas fluorescentes adecuadas incluyen proteínas verdes fluorescentes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), proteínas amarillas fluorescentes (por ejemplo, YFP, EYFP, Citrina, Venus, YPet, PhiYFP, ZsYellow1.), proteínas azules fluorescentes (por ejemplo EBFP, EBFP2, Azurita, mKalama1, GFPuv, Sapphire, T-sapphire), proteínas cian fluorescentes (por ejemplo ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas rojas fluorescentes (mKate, mKate2, mPlum, monómero DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tándem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), y proteínas naranjas fluorescentes (mOrange, mKO, Kusabira-Orange, Kusabira-Orange monomérica, 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 epítipo. Las etiquetas ejemplares incluyen, aunque no de forma limitativa, glutatión-S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tioredoxina (TRX), poli(NANP), etiqueta de purificación por afinidad en tándem (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 transportadora de carboxil biotina (BCCP) y calmodulina.

En determinados ejemplos, la endonucleasa guiada por ARN puede ser parte de un complejo proteína-ARN que comprende un ARN guía. El ARN guía interacciona con la endonucleasa guiada por ARN para dirigir la endonucleasa a un sitio diana específico, en el que el extremo 5' del ARN guía alinea sus bases con una secuencia de protoespaciador específica.

(II) Ácidos nucleicos que codifican endonucleasas guiadas por ARN o proteínas de fusión

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 o la proteína de fusión es ARNm. El ARNm puede tener casquete en 5' y/o estar poliadenilado en 3'. En otra realización, el ácido nucleico que codifica la endonucleasa guiada por ARN o la proteína de fusión es ADN. El ADN puede estar presente en un vector (véase a continuación).

El ácido nucleico que codifica la endonucleasa guiada por ARN o la proteína de fusión puede tener un codón optimizado para una traducción eficaz a proteína en la célula eucariota o animal de interés. Por ejemplo, se pueden optimizar codones para la expresión en seres humanos, ratones, ratas, hámsteres, vacas, cerdos, gatos, perros, peces, anfibios, plantas, levadura, insectos, etcétera (véase la base de datos del uso de codones en www.kazusa.or.jp/codon/). Los programas para la optimización de codones están disponibles como programas de dominio público (por ejemplo, OPTIMIZER en genomes.urv.es/OPTIMIZER; OptimumGene™ de GenScript en www.genscript.com/codon_opt.html). Los programas comerciales de optimización de codones también están disponibles.

El ADN que codifica la endonucleasa guiada por ARN se puede enlazar de manera operativa con al menos una secuencia de control del promotor. En algunas repeticiones, la secuencia codificante de ADN se puede unir de manera operativa a una secuencia de control del promotor para la expresión en la célula eucariota o animal de interés. La secuencia de control del promotor puede ser constitutiva, regulada o específica de tejido. Las secuencias de control del promotor constitutivas adecuadas incluyen, pero sin limitación, el promotor temprano inmediato de citomegalovirus (CMV), el promotor del virus del simio (SV40), el promotor tardío principal de adenovirus, el promotor del virus del sarcoma de Rous (VSR), el promotor del virus del tumor mamario del ratón (MMTV), el promotor de la

fosfoglicerato cinasa (PGK), el promotor del factor de elongación (ED1)-alfa, los promotores de ubiquitina, los promotores de actina, los promotores de tubulina, los promotores de inmunoglobulina, fragmentos de los mismos o combinaciones de cualquiera de los anteriores. Los ejemplos de secuencias de control del promotor reguladas incluyen sin limitación aquellas reguladas por choque térmico, metales, esteroides, antibióticos o alcohol. Los ejemplos no limitantes de promotores específicos de tejido incluyen el promotor B29, el promotor de CD14, el promotor de CD43, el promotor de CD45, el promotor de CD68, el promotor de desmina, el promotor de elastasa-1, el promotor de endoglina, el promotor de fibronectina, el promotor de Flt-1, el promotor de GFAP, el promotor de GPIIb, el promotor de ICAM-2, el promotor de INF- β , el promotor de Mb, el promotor de Nphsl, el promotor de OG-2, el promotor de SP-B, el promotor de SYN1 y el promotor de WASP. La secuencia del promotor puede ser de tipo silvestre o se puede modificar para una expresión más eficiente o eficaz. En un ejemplo, el ADN codificante puede unirse de manera operativa a un promotor de CMV para la expresión constitutiva en células de mamífero.

La secuencia que codifica la endonucleasa guiada por ARN puede estar unida de manera operativa a una secuencia de promotor que es reconocida por una ARN polimerasa del fago para la síntesis *in vitro* de ARNm. En tal ejemplo, el ARN transcrito *in vitro* se puede purificar para su uso en los procedimientos detallados anteriormente en las secciones (III) y (IV). Por ejemplo, la secuencia del promotor puede ser una secuencia del promotor T7, T3 o SP6 o una variación de la secuencia del promotor T7, T3 o SP6. En una realización a modo de ejemplo, el ADN que codifica la proteína está unido de manera operativa a un promotor T7 para la síntesis *in vitro* de ARNm usando una ARN polimerasa de T7.

En una alternativa, la secuencia que codifica la endonucleasa guiada por ARN puede estar unida de manera operativa a una secuencia del promotor para la expresión *in vitro* de la endonucleasa guiada por ARN en células bacterianas o eucariotas. En esos casos, la proteína expresada se puede purificar para su uso en los procedimientos detallados a continuación en las secciones (III) y (IV). Los promotores bacterianos adecuados incluyen, sin límite, promotores T7, promotores del operón lac, promotores *trp*, variaciones de los mismos, y combinaciones de los mismos. Un promotor bacteriano ejemplar es *tac*, que es un híbrido de los promotores *trp* y *lac*. Los ejemplos no limitantes de los promotores eucarióticos se enumeran anteriormente.

En aspectos adicionales, el ADN que codifica la endonucleasa guiada por ARN o la proteína de fusión también puede estar unida a una señal de poliadenilación (por ejemplo, señal poliA del SV40, señal poliA de la hormona de crecimiento bovina (BGH), etc.) y/o al menos una secuencia de terminación transcripcional. Además, la secuencia que codifica la endonucleasa guiada por ARN o la proteína de fusión también puede estar unida 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 se detallan anteriormente en la sección (I).

En diversas realizaciones, el ADN que codifica la endonucleasa guiada por ARN o la proteína de fusión puede estar presente en un vector. Los vectores adecuados incluyen vectores de plásmidos, fagémidos, cósmidos, minicromosomas artificiales, transposones y vectores víricos (por ejemplo, vectores lentivíricos, vectores víricos adenoasociados, etc.). En un ejemplo, el ADN que codifica la endonucleasa guiada por ARN o la proteína de fusión está presente en un vector de plásmido. Los ejemplos no limitantes de vectores de plásmidos adecuados incluyen pUC, pBR322, pET, pBluescript, y variantes de los mismos. El vector puede comprender secuencias adicionales de control de la expresión (por ejemplo, secuencias potenciadoras, secuencias de Kozak, secuencias de poliadenilación, secuencias de terminación transcripcional, etc.), secuencias de marcador seleccionable (por ejemplo, genes de resistencia a antibióticos), orígenes de replicación y similares. La información adicional se puede encontrar en Current Protocols in Molecular Biology" de Ausubel y col., John Wiley & Sons, Nueva York, 2003 o "Molecular Cloning: A Laboratory Manual" de Sambrook y Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3ª edición, 2001.

En algunos ejemplos, el vector de expresión que comprende la secuencia que codifica la endonucleasa guiada por ARN o la proteína de fusión puede comprender adicionalmente la secuencia que codifica un ARN guía. La secuencia que codifica el ARN guía generalmente está unido de manera operativa a al menos una secuencia de control de la transcripción para la expresión del ARN guía en la célula o embrión de interés. Por ejemplo, el ADN que codifica el ARN guía se puede unir de manera operativa a una secuencia del promotor que se reconoce mediante la ARN polimerasa III (Pol III). Los ejemplos de promotores de Pol III incluyen, pero sin limitación, los promotores de ARN U6, U3, H1 y 7SL de mamífero.

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

Como se ha indicado anteriormente, la presente invención abarca un procedimiento para modificar una secuencia cromosómica en una célula eucariota, comprendiendo el procedimiento introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas y, opcionalmente, un polinucleótido donador, en el que cada sistema de nickasa guiado por ARN comprende (i) una endonucleasa guiada por ARN que está modificada para escindir una cadena de una secuencia de cadena doble y (ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN, en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM), y los sitios de diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y cultivar la célula eucariota de manera que las dos

endonucleasas guiadas por ARN escinden cadenas opuestas de la secuencia cromosómica en estrecha cercanía, lo suficiente como para inducir una rotura de cadena doble en la secuencia cromosómica, y reparar la rotura de cadena doble mediante un proceso de reparación de ADN que lleva a la modificación de la secuencia cromosómica, en el que el procedimiento no comprende un proceso de 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.

La invención también proporciona un procedimiento ex vivo o in vitro de modificar una secuencia exógena en una secuencia cromosómica en una célula eucariota, comprendiendo el procedimiento:

(a) introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas y, opcionalmente, un polinucleótido donador, en los que los sistemas de nickasa guiados por ARN comprenden

(i) una nucleasa guiada por ARN que está modificada para escindir una cadena de una secuencia de cadena doble; y

(ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN,

en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM) y los sitios diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y

b) cultivar la célula eucariota de tal manera que las dos endonucleasas guiadas por ARN escinden cadenas opuestas de la secuencia cromosómica en estrecha cercanía, lo suficiente para introducir una rotura de cadena doble en la secuencia cromosómica, y reparar la rotura de cadena doble por un proceso de reparación de ADN que lleva a una modificación de la secuencia cromosómica, en la que el procedimiento no comprende un proceso para modificar la identidad genética de la línea germinal de un ser humano.

En algunas realizaciones, el procedimiento puede comprender introducir una endonucleasa guiada por ARN modificada para escindir una cadena de una secuencia de cadena doble (o ácido nucleico codificante) y dos ARN guía (o ADN codificante) en una célula o embrión, en la que cada ARN guía dirige la endonucleasa guiada por ARN hacia un sitio diana específico, el sitio en el que la endonucleasa modificada escinde una cadena (es decir, hace una incisión) de la secuencia cromosómica de cadena doble, y en la que las dos incisiones están en cadenas opuestas y en estrecha cercanía, lo suficiente para constituir una rotura de cadena doble. Véase la FIG. 1. En realizaciones en las que el polinucleótido donador opcional no está presente, la rotura de cadena doble se puede reparar por un proceso de reparación no homóloga de manera que se puedan dar durante el proceso de reparación de la rotura delecciones de al menos un nucleótido, inserciones de al menos un nucleótido, sustituciones de al menos un nucleótido o combinaciones de las mismas. En realizaciones en las que el polinucleótido donador opcional está presente, la secuencia donadora en el polinucleótido donador se puede intercambiar con o integrar en la secuencia cromosómica durante la reparación de las roturas de doble cadena bien mediante un proceso de reparación basado en homología (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada aguas arriba y aguas abajo por secuencias que tienen una identidad de secuencia sustancial con las secuencias de aguas arriba y aguas abajo, respectivamente, de los sitios dirigidos en la secuencia cromosómica) o bien un proceso de reparación no homóloga (por ejemplo, en realizaciones en las que la secuencia donadora está flanqueada por proyecciones compatibles).

(a) Endonucleasa guiada por ARN

El procedimiento comprende introducir en una célula al menos una endonucleasa guiada por ARN que comprende al menos una señal de localización nuclear o ácido nucleico que codifica al menos una endonucleasa guiada por ARN que comprende al menos una señal de localización nuclear. Tales 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 procedimientos excluyen los que comprenden un proceso para modificar la identidad genética de la línea germinal de un ser humano.

En algunas realizaciones, la endonucleasa guiada por ARN se puede introducir en la célula o embrión como una proteína aislada. En dichas realizaciones, la endonucleasa guiada por ARN puede comprender adicionalmente al menos un dominio de penetración celular, que facilita la captación celular de la proteína. En otras realizaciones, la endonucleasa guiada por ARN se puede introducir en la célula o embrión como una molécula de ARNm. Aún en otras realizaciones, la endonucleasa guiada por ARN se puede introducir en la célula o embrión como una molécula de ADN. En general, la secuencia de ADN que codifica la proteína de fusión está unida de manera operativa a una secuencia del promotor 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. Aún en otras realizaciones, se puede introducir la proteína de fusión en la célula o embrión como un complejo ARN-proteína que comprende la proteína de fusión y el ARN guía.

En realizaciones alternativas, el ADN que codifica la endonucleasa guiada por ARN puede comprender

adicionalmente la secuencia que codifica un ARN guía. En general, cada una de las secuencias que codifican la endonucleasa guiada por ARN y el ARN guía están unidas de manera operativa la secuencia de control del promotor apropiada que permite la expresión de la endonucleasa guiada por ARN y el ARN guía, respectivamente, en la célula o embrión. La secuencia de ADN que codifica la endonucleasa guiada por ARN y el ARN guía puede comprender además secuencia(s) de control de la expresión, reguladoras y/o de procesamiento. 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 la introducción en una célula de al menos un ARN guía o ADN que codifica al menos un ARN guía. Un ARN guía interacciona con la endonucleasa guiada por ARN para dirigir la endonucleasa a un sitio diana específico, en cuyo sitio, el extremo 5' del ARN guía alinea sus bases con una secuencia de protoespaciador específica en la secuencia cromosómica.

Cada ARN guía comprende tres regiones: una primera región en el extremo 5' que es complementaria con el sitio diana en la secuencia cromosómica, una segunda región interna que forma una estructura en tallo-bucle, y una tercera región 3' que esencialmente permanece monocatenaria. La primera región de cada ARN guía es diferente, de manera que cada ARN guía a una proteína de fusión hacia 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 con la secuencia (es decir, la secuencia del protoespaciador) en el sitio diana en la secuencia cromosómica de manera que la primera región del ARN guía puede alinear sus bases con el sitio dirigido. En diversas realizaciones, la primera región del ARN guía puede comprender desde aproximadamente 10 nucleótidos a más de aproximadamente 25 nucleótidos. Por ejemplo, la región del alineamiento 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 a modo de ejemplo, la primera región de ARN guía es de 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 (u horquilla) y un bucle. La longitud del bucle y del tallo puede variar. Por ejemplo, el bucle puede variar desde aproximadamente 3 hasta aproximadamente 10 nucleótidos de longitud, y el tallo puede variar desde aproximadamente 6 hasta aproximadamente 20 pares de bases de longitud. El tallo puede comprender una o más protuberancias de 1 a aproximadamente 10 nucleótidos. De este modo, la longitud global de la segunda región puede variar desde aproximadamente 16 hasta aproximadamente 60 nucleótidos de longitud. En una realización ejemplar, el bucle es de 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 monocatenario. De este modo, 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 es más de aproximadamente 4 nucleótidos de longitud. Por ejemplo, la longitud de la tercera región puede variar desde aproximadamente 5 hasta aproximadamente 60 nucleótidos de longitud.

La longitud combinada de la segunda y tercera región (también llamada la región universal o estructural) del ARN guía puede variar desde aproximadamente 30 hasta aproximadamente 120 nucleótidos de longitud. En un aspecto, la longitud combinada de la segunda y tercera región del ARN guía varía desde aproximadamente 70 hasta aproximadamente 100 nucleótidos de longitud.

En algunas realizaciones, el ARN guía comprende una única molécula que comprende las tres regiones. En otras realizaciones, el ARN guía puede comprender dos moléculas separadas. La primera molécula de ARN puede comprender la primera región del ARN guía y una mitad del "tallo" de la segunda región del ARN guía. La segunda molécula de ARN puede comprender la otra mitad del "tallo" de la segunda región del ARN guía y la tercera región del ARN guía. De este modo, en esta realización, la primera y segunda molécula de ARN contiene cada una una secuencia de nucleótidos que son complementarias entre sí. Por ejemplo, en una realización, la primera y segunda molécula de ARN comprende cada una una secuencia (de aproximadamente 6 a aproximadamente 20 nucleótidos) que alinea sus bases con la otra secuencia para formar un ARN guía funcional.

En algunas realizaciones, el ARN guía se puede introducir en la célula o embrión como una molécula de ARN. La molécula de ARN se puede transcribir *in vitro*. Como alternativa, la molécula de ARN se puede sintetizar químicamente.

En otras realizaciones, el ARN guía se puede introducir en la célula o embrión como una molécula de ADN. En esos casos, el ADN que codifica el ARN guía se puede unir de manera operativa a la secuencia de control del promotor para la expresión del ARN guía en la célula o embrión de interés. Por ejemplo, la secuencia codificante de ARN se puede unir de manera operativa a una secuencia del promotor que se reconoce mediante la ARN polimerasa III (Pol III). Los ejemplos de promotores de Pol III incluyen, pero sin limitación, promotores U6 o H1 de mamífero. En una realización ejemplar, la secuencia codificante de ARN se enlaza a un promotor U6 humano o de ratón. En otras

realizaciones ejemplares, la secuencia codificante de ARN se enlaza a un promotor H1 humano o de ratón.

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 de plásmidos, fagémidos, cósmidos, minicromosomas artificiales, transposones y vectores víricos. En una realización a modo de ejemplo, el ADN que codifica la endonucleasa guiada por ARN está presente en un vector de plásmido. Los ejemplos no limitantes de vectores de plásmidos adecuados incluyen pUC, pBR322, pET, pBluescript, y variantes de los mismos. El vector puede comprender secuencias adicionales de control de la expresión (por ejemplo, secuencias potenciadoras, secuencias de Kozak, secuencias de poliadenilación, secuencias de terminación transcripcional, etc.), secuencias de marcador seleccionable (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 secuencia codificante de proteína y un segundo vector que contiene la secuencia codificante de ARN guía) o ambas pueden ser parte de la misma molécula (por ejemplo, un vector que contiene la secuencia codificante (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 rotura de doble cadena en la secuencia cromosómica. El sitio diana no tiene limitación de secuencia excepto en que la misma secuencia va seguida de manera inmediata (aguas abajo) de una secuencia consenso. Esta secuencia consenso también se conoce como motivo adyacente al protoespaciador (PAM, del inglés *protospacer adjacent motif*). Los ejemplos de PAM incluyen, pero sin limitación, NGG, NGGNG y NNAGAAW (en los que N se define como cualquier nucleótido y W se define bien como A o como T). Tal como se detalla anteriormente en la sección (IV)(b), la primera región (en el extremo 5') del ARN guía es complementaria con el protoespaciador de la secuencia diana. Normalmente, la primera región del ARN guía es de aproximadamente 19 a 21 nucleótidos de longitud. De este modo, en determinados aspectos, la secuencia del sitio diana en la secuencia cromosómica es 5'-N₁₉₋₂₁-NGG-3'. El PAM está en letras cursivas.

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 proteína o un gen codificante de ARN. El gen puede ser cualquier gen de interés.

(d) Polinucleótido donador opcional

En algunas realizaciones, el procedimiento comprende adicionalmente introducir al menos un polinucleótido donador en la célula. Un polinucleótido donador comprende al menos una secuencia donadora. En algunos aspectos, una secuencia del polinucleótido donador se corresponde con la secuencia cromosómica natural. Por ejemplo, la secuencia donadora puede ser esencialmente idéntica a una parte de la secuencia cromosómica en o cerca del sitio direccionado, pero que comprende al menos un cambio de nucleótido. De este modo, la secuencia donadora puede comprender una versión modificada de la secuencia de tipo silvestre en el sitio direccionado de manera que, tras la integración o intercambio con la secuencia natural, la secuencia en la localización cromosómica direccionada comprende al menos un cambio de nucleótido. Por ejemplo, el cambio puede ser una inserción de uno o más nucleótidos, una delección de uno o más nucleótidos, una sustitución de uno o más nucleótidos, o combinaciones de los mismos. Como consecuencia de la integración de la secuencia modificada, la célula o el embrión/animal puede producir un producto génico modificado de la secuencia cromosómica direccionada.

En otros aspectos, la secuencia donadora del polinucleótido donador se corresponde con una secuencia exógena. Tal como se usa en el presente documento, una secuencia "exógena" se refiere a una secuencia que no es natural para la célula o embrión, o una secuencia cuya localización natural en el genoma de la célula o embrión está en una localización diferente. Por ejemplo, la secuencia exógena puede comprender secuencia codificante de proteína, que puede estar unida de manera operativa a una secuencia de control del promotor 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 se puede integrar en la secuencia cromosómica de manera que su expresión se regule mediante una secuencia de control del promotor endógena. En otras repeticiones, la secuencia exógena puede ser una secuencia de control transcripcional, otra secuencia de control de la expresión, una secuencia codificante de ARN, etcétera. La integración de una secuencia exógena en una secuencia cromosómica se denomina "inserción".

Como se apreciará por los expertos en la materia, la longitud de la secuencia donadora puede variar y lo hará. Por ejemplo, la secuencia donadora puede variar en longitud desde varios nucleótidos hasta cientos de miles de nucleótidos.

Polinucleótido donador que comprende secuencias aguas arriba y aguas abajo. En algunas realizaciones, la secuencia donadora en el polinucleótido donador está flanqueado por una secuencia aguas arriba y una secuencia

aguas abajo, que tiene identidad de secuencia sustancial con secuencias localizadas aguas arriba y aguas abajo, respectivamente, del sitio dirigido en la secuencia cromosómica. Debido a estas similitudes de secuencia, las secuencias aguas arriba y aguas abajo del polinucleótido donador permiten la recombinación homóloga entre el polinucleótido donador y la secuencia cromosómica dirigida de manera que la secuencia del donador se puede integrar en (o intercambiar con) la secuencia cromosómica.

La secuencia aguas arriba, tal como se usa en el presente documento, se refiere a una secuencia de ácido nucleico que comparte la identidad de secuencia sustancial con una secuencia cromosómica aguas arriba del sitio dirigido. De forma análoga, la secuencia aguas abajo se refiere a una secuencia de ácido nucleico que comparte la identidad de secuencia sustancial con una secuencia cromosómica aguas abajo del sitio dirigido. Tal como se usa en el presente documento, la frase "identidad de secuencia sustancial" se refiere a secuencias que tienen al menos aproximadamente el 75 % de identidad de secuencia. De este modo, las secuencias aguas arriba y aguas abajo en el polinucleótido donador pueden tener aproximadamente el 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con la secuencia aguas arriba o aguas abajo del sitio dirigido. En una realización ejemplar, las secuencias aguas arriba y aguas abajo en el polinucleótido donador pueden tener aproximadamente el 95 % o el 100 % de identidad de secuencia con las secuencias cromosómicas aguas arriba o aguas abajo del sitio dirigido. En una realización, la secuencia aguas arriba comparte la identidad de secuencia sustancial con una secuencia cromosómica localizada inmediatamente aguas arriba del sitio dirigido (es decir, adyacente al sitio dirigido). En otras realizaciones, la secuencia aguas arriba comparte la identidad de secuencia sustancial con una secuencia cromosómica que se localizan en aproximadamente cien (100) nucleótidos aguas arriba del sitio dirigido. Por lo tanto, por ejemplo, la secuencia de aguas arriba puede compartir identidad de secuencia sustancial con una secuencia cromosómica que se localiza 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 aguas arriba desde el sitio dirigido. En una realización, la secuencia aguas abajo comparte la identidad de secuencia sustancial con una secuencia cromosómica localizada inmediatamente aguas abajo del sitio dirigido (es decir, adyacente al sitio dirigido). En otras realizaciones, la secuencia aguas abajo comparte la identidad de secuencia sustancial con una secuencia cromosómica que se localizan en aproximadamente cien (100) nucleótidos aguas abajo del sitio dirigido. Por lo tanto, por ejemplo, la secuencia de aguas abajo puede compartir identidad de secuencia sustancial con una secuencia cromosómica que se localiza 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 aguas abajo desde el sitio dirigido.

Cada secuencia aguas arriba o aguas abajo puede variar en longitud desde aproximadamente 20 nucleótidos a aproximadamente 5000 nucleótidos. En algunas realizaciones, las secuencias aguas arriba y aguas 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 una realización ejemplar, las secuencias aguas arriba o aguas abajo pueden variar en longitud desde aproximadamente 50 a aproximadamente 1500 nucleótidos.

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

Polinucleótidos donadores que comprenden sitio(s) de escisión dirigido(s). En otras realizaciones, el polinucleótido donador puede comprender adicionalmente al menos un sitio de escisión dirigido que se reconoce mediante la endonucleasa guiada por ARN. El sitio de escisión dirigido añadido al polinucleótido donador se puede colocar aguas arriba o aguas abajo o ambos, aguas arriba y aguas abajo de la secuencia donadora. Por ejemplo, la secuencia donadora puede estar flanqueada mediante sitios de escisión dirigidos de manera que, tras la escisión por la endonucleasa guiada por ARN, la secuencia donadora está flanqueada por proyecciones que son compatibles con las de la secuencia cromosómica generadas tras la escisión mediante la endonucleasa guiada por ARN. Por consiguiente, la secuencia donadora se puede unir con la secuencia cromosómica escindida durante la reparación de la rotura de doble cadena mediante un proceso de reparación no homóloga. Generalmente, los polinucleótidos donadores que comprenden el(los) sitio(s) de escisión serán circulares (por ejemplo, pueden ser parte de un vector de plásmido).

Polinucleótido donador que comprende una secuencia donadora corta con proyecciones opcionales. En más realizaciones alternativas, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta con proyecciones cortas opcionales que son compatibles con las proyecciones generadas mediante la endonucleasa guiada por ARN. En dichas realizaciones, la secuencia donadora se puede unir directamente con la secuencia cromosómica escindida durante la reparación de la rotura de doble cadena. En algunos casos, la secuencia donadora puede ser menor de aproximadamente 1.000, menor de aproximadamente 500, menor de aproximadamente 250, o menor de aproximadamente 100 nucleótidos. En determinados casos, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta con extremos romos. En otras repeticiones, el polinucleótido donador puede ser una molécula lineal que comprende una secuencia donadora corta

con proyecciones en 5' y/o 3'. Las proyecciones pueden comprender 1, 2, 3, 4 o 5 nucleótidos.

Normalmente, el polinucleótido donador será ADN. El ADN puede ser monocatenario o bicatenario y/o lineal o circular. El polinucleótido donador puede ser un plásmido de ADN, un cromosoma artificial bacteriano (BAC, del inglés *bacterial artificial chromosome*), un cromosoma artificial de levadura (YAC, del inglés *yeast artificial chromosome*), un vector vírico, una parte lineal de ADN, un fragmento de PCR, un ácido nucleico desnudo, o un ácido nucleico que forma complejo con un vehículo de administración tal como un liposoma o un poloxámero. En determinadas realizaciones, el polinucleótido donador que comprende la secuencia donadora puede ser parte de un vector de plásmido. En cualquiera de estas situaciones, el polinucleótido donador que comprende la secuencia donadora puede comprender adicionalmente al menos una secuencia adicional.

(e) Introducción en la célula

La(s) endonucleasa(s) dirigida(s) por ARN (o el ácido nucleico codificante), el(los) ARN guía(s) (o ADN codificante) y el(los) polinucleótido(s) donador(es) opcional(es) se pueden introducir en una célula mediante una variedad de medios. En algunas realizaciones, se transfecta la célula. Los procedimientos de transfección adecuados incluyen la transfección mediada por fosfato cálcico, la nucleofección (o electroporación), la transfección por polímeros catiónicos (por ejemplo, DEAE-dextrano o polietilénimina), la transducción vírica, la transfección por virosoma, la transfección por virión, la transfección por liposoma, la transfección por liposoma catiónico, la transfección por inmunoliposoma, la transfección por lípido no liposoma, la transfección por dendrímero, la transfección por choque térmico, la magnetofección, la lipofección, la biobalística, la impalefacción, la sonoporación, la transfección óptica y la captación de ácidos nucleicos potenciada por un agente patentado. Los procedimientos de transfección son bien conocidos en la materia (véase, por ejemplo, "Current Protocols in Molecular Biology" Ausubel y col., John Wiley & Sons, Nueva York, 2003 o "Molecular Cloning: A Laboratory Manual" de Sambrook y Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3ª Edición, (2001)). En otras realizaciones, las moléculas se introducen en la célula mediante microinyección. Normalmente, cuando la célula es una célula célula embrionaria de animal no humano, el embrión es un embrión de etapa unicelular fecundado de la especie de interés. Por ejemplo, las moléculas se pueden inyectar en los pronúcleos de embriones unicelulares.

La(s) endonucleasa(s) dirigida(s) por ARN (o el ácido nucleico codificante), el(los) ARN guía(s) (o los ADN que codifican el ARN guía) y el(los) polinucleótido(s) donador(es) opcional(es) se pueden introducir en la célula de manera simultánea o de manera secuencial. La proporción de endonucleasa(s) dirigida(s) por ARN (o ácido nucleico codificante) frente al(los) ARN guía (o ADN codificante), generalmente será de aproximadamente una estequiometría tal 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 administran juntos en el vector de plásmido.

(f) Cultivo de la célula

El procedimiento además comprende el mantenimiento de la célula o embrión en condiciones apropiadas, de manera que el(los) ARN guía dirija a la(s) endonucleasa(s) guiada(s) por ARN al(los) sitios dirigido(s) en la secuencia cromosómica, y la(s) endonucleasa(s) guiada por ARN introduzca(n) al menos una rotura de doble cadena en la secuencia cromosómica. Una rotura de doble cadena se puede reparar mediante un proceso de reparación de ADN de manera que la secuencia cromosómica se modifique mediante una 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.

En realizaciones en las que no se introducen polinucleótidos donadores en la célula o embrión, la rotura de doble cadena podría repararse mediante un proceso de reparación por unión de extremos no homólogos (NHEJ, del inglés *non-homologous end-joining*). Dado que la NHEJ es propensa a errores, las delecciones de al menos un nucleótido, las inserciones de al menos un nucleótido, las sustituciones de al menos un nucleótido o las combinaciones de las mismas pueden tener lugar durante la reparación de la rotura. Por consiguiente, la secuencia en la secuencia cromosómica se puede modificar de manera que el marco de lectura de una región codificante se pueda desplazar y la secuencia cromosómica se inactíve o se "suprima". Una secuencia cromosómica inactivada que codifica una proteína no produce la proteína codificada por la secuencia cromosómica de tipo silvestre.

En realizaciones en las que un polinucleótido donador que comprende secuencias aguas arriba y aguas abajo se introduce en la célula o embrión, la rotura de doble cadena se puede reparar mediante un proceso de reparación dirigida por homología (HDR, del inglés *homology-directed repair*) de manera que la secuencia donadora se integra en la secuencia cromosómica. Por consiguiente, se puede integrar una secuencia exógena en el genoma de la célula o embrión, o la secuencia cromosómica dirigida se puede modificar mediante el intercambio de una secuencia modificada por la secuencia cromosómica de tipo silvestre.

En realizaciones en las que un polinucleótido donador que comprende el sitio de escisión se introduce en la célula o embrión, la endonucleasa guiada por ARN puede escindir tanto la secuencia cromosómica dirigida como el polinucleótido donador. El polinucleótido donador linealizado se puede integrar en la secuencia cromosómica en el sitio de la rotura de doble cadena mediante unión entre el polinucleótido donador y la secuencia cromosómica escindida mediante un proceso de NHEJ.

En realizaciones en las que un polinucleótido donador lineal que comprende una secuencia donadora corta se introduce en la célula o embrión, la secuencia donadora corta se puede integrar en la secuencia cromosómica en el sitio de la rotura de doble cadena mediante un proceso de NHEJ. La integración puede tener lugar mediante la unión de extremos romos entre la secuencia donadora corta y la secuencia cromosómica en el sitio de la rotura de doble cadena. Como alternativa, la integración puede tener lugar mediante la unión de los extremos cohesivos (es decir, que tienen proyecciones en 5' o en 3') entre una secuencia donadora corta que está flanqueada por proyecciones que son compatibles con las generadas mediante la endonucleasa de direccionamiento de ARN en la secuencia cromosómica escindida.

En general, la célula se mantiene en condiciones apropiadas para el crecimiento celular y/o su conservación. Las condiciones adecuadas de cultivo celular son bien conocidas en la materia 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 el cultivo de células son conocidos en la materia y pueden variar y lo harán dependiendo del tipo celular. Se puede usar la optimización habitual, en todos los casos, para determinar las mejores técnicas para un tipo celular en particular.

Se puede cultivar un embrión *in vitro* (por ejemplo, en cultivo celular). Normalmente, el embrión se cultiva a una temperatura apropiada y en medios apropiados con la proporción de O₂/CO₂ necesaria como para permitir la expresión de la endonucleasa guiada por ARN y del ARN guía, si fuese necesario. Los ejemplos no limitantes adecuados de medios incluyen los medios M2, M16, KSOM, BIOC y HTF. Un experto en la materia apreciará que las condiciones del cultivo pueden variar y lo harán dependiendo de la especie del embrión. Se puede usar la optimización habitual, en todos los casos, para determinar las mejores condiciones de cultivo para una especie particular de embrión. En algunos casos, una línea celular puede provenir de un embrión cultivado *in vitro* (por ejemplo, una línea de células madre embrionarias).

Como alternativa, se puede cultivar un embrión *in vivo* transfiriendo el embrión al útero de un hospedador femenino. Hablando de manera general, el hospedador femenino es de la misma especie o similar que la del embrión. Preferentemente, el hospedador femenino está pseudoembarazado. Los procedimientos para preparar hospedadores femeninos pseudoembarazados se conocen en la materia. Además, los procedimientos para transferir un embrión en un hospedador femenino son conocidos. El cultivo de un embrión *in vivo* permite el desarrollo del embrión y pueden dar como resultado un nacimiento vivo de un animal que proviene del embrión. Tal animal comprendería la secuencia cromosómica modificada en cada célula del cuerpo. Se excluyen de manera específica del ámbito de la invención los procedimientos que comprenden modificar la identidad genética de la línea germinal de un ser humano.

(g) Tipos de células y embriones

Una variedad de células eucariotas y embriones son adecuados para su uso en el procedimiento. 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 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 ejemplares, incluyendo los embriones unicelulares, incluyen sin limitación los embriones de ratón, de rata, de hámster, de roedor, de conejo, de gato, de perro, de oveja, de cerdo, de vaca, de caballo y de primate. Aún en otras realizaciones, la célula puede ser una célula madre. Las células madre adecuadas incluyen sin limitación las células madre embrionarias, las células madre de tipo ES, células madre fetales, células madre de adulto, células madre pluripotenciales, células madre de pluripotencialidad inducida, células madre multipotentes, células madre oligopotentes, células madre unipotentes y otras. En realizaciones ejemplares, la célula es una célula de mamífero.

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

(ATCC, Manassas, VA).

(IV) Células y animales modificados genéticamente

La presente divulgación abarca células embriones no humanos y animales no humanos modificados genéticamente, que comprenden al menos una secuencia cromosómica que se ha modificado usando el proceso mediado por endonucleasa guiada por ARN, usando los procedimientos descritos en el presente documento. La divulgación proporciona células que comprenden al menos una molécula de ADN o de ARN que codifica una endonucleasa guiada por ARN dirigida hacia una secuencia cromosómica de interés, al menos un ARN guía y, opcionalmente, uno o más polinucleótidos donadores. La divulgación también proporciona embriones no humanos que comprenden al menos una molécula de ADN o de ARN que codifica una endonucleasa guiada por ARN o proteína de fusión dirigida hacia una secuencia cromosómica de interés, al menos un ARN guía y opcionalmente uno o más polinucleótidos donadores.

La presente divulgación proporciona animales no humanos, embriones no humanos o células animales modificados genéticamente que comprenden al menos una secuencia cromosómica modificada. La secuencia cromosómica modificada se puede modificar de manera que (1) sea inactiva, (2) tenga una expresión alterada o produzca un producto de proteína alterada o (3) comprenda una secuencia integrada. La secuencia cromosómica se modifica con un proceso mediado por endonucleasa guiada por ARN, usando los procedimientos descritos en el presente documento.

Tal como se ha tratado, un aspecto de la presente divulgación proporciona un animal modificado genéticamente 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 se puede inactivar de manera que la secuencia no se transcriba y/o no se produzca una proteína funcional. Por lo tanto, un animal modificado genéticamente que comprende una secuencia cromosómica inactivada puede denominarse un "animal genomanipulado por supresión génica" o un "animal genomanipulado secundariamente por supresión génica". 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 interruptora (es decir, sustitución de un único nucleótido por otro nucleótido de manera que se introduzca un codón de parada). Como consecuencia de la mutación, la secuencia cromosómica dirigida se inactiva y no se produce una proteína funcional. La secuencia cromosómica inactivada no comprende una secuencia introducida de manera exógena. También se incluyen en el presente documento animales modificados genéticamente 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 se puede alterar de manera que codifique un producto variante de proteína. Por ejemplo, un animal modificado genéticamente que comprende una secuencia cromosómica modificada puede comprender una(s) mutación(es) puntual(es) dirigidas u otra modificación de manera que se produzca un producto de proteína alterada. En una realización, la secuencia cromosómica se puede modificar de manera que al menos un nucleótido se cambie y la proteína expresada comprenda un resto de aminoácido cambiado (mutación de aminoácido). En otra realización, la secuencia cromosómica se puede modificar para que comprenda más de una mutación de aminoácido de manera que se cambie más de un aminoácido. Además, la secuencia cromosómica se puede modificar para que tenga una delección o inserción de tres nucleótidos de manera que la proteína expresada comprenda una única delección o inserción de aminoácido. La proteína variante o alterada puede tener propiedades o actividades alteradas en comparación con la proteína de tipo silvestre, tal como una especificidad por el sustrato alterada, una actividad enzimática alterada, unas tasas cinéticas alteradas, etc.

En otra realización, el animal modificado genéticamente puede comprender al menos una secuencia integrada en el cromosoma. Un animal modificado genéticamente que comprende una secuencia integrada puede denominarse un "animal genomanipulado por inserción génica" o un "animal genomanipulado secundariamente por inserción génica". La secuencia integrada en el cromosoma 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 se puede integrar en una secuencia cromosómica que codifica una proteína de manera que la secuencia cromosómica se inactiva, pero la secuencia exógena se expresa. En tal caso, la secuencia que codifica la proteína ortóloga o la proteína endógena puede estar unida de manera operativa a una secuencia de control del promotor. Como alternativa, una secuencia que codifica una proteína ortóloga o una proteína endógena se puede integrar en una secuencia cromosómica sin afectar a la expresión de una secuencia cromosómica. Por ejemplo, una secuencia que codifica una proteína se puede integrar en un locus "de sitio seguro", tal como el locus Rosa26, el locus HPRT, o el locus AAV. La presente divulgación también abarca animales modificados genéticamente en los que dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez o más secuencias, incluyendo secuencias que codifican proteína(s), se integran en el genoma.

La secuencia integrada en el cromosoma que codifica una proteína puede codificar la forma de tipo silvestre de una proteína de interés o puede codificar una proteína que comprende al menos una modificación de manera que se produce una versión alterada de la proteína. Por ejemplo, una secuencia integrada en el cromosoma que codifica una proteína relacionada con una enfermedad o trastorno puede comprender al menos una modificación de manera que la versión alterada de la proteína producida provoque o potencie el trastorno asociado. Como alternativa, la

secuencia integrada en el cromosoma que codifica una proteína relacionada con una enfermedad o trastorno puede comprender al menos una modificación de manera que la versión alterada de la proteína proteja frente al desarrollo del trastorno asociado.

En un ejemplo adicional, el animal modificado genéticamente puede ser un animal "humanizado" que comprende al menos una secuencia integrada en el cromosoma que codifica una proteína humana funcional. La proteína humana funcional puede no tener el correspondiente ortólogo en el animal modificado genéticamente. Como alternativa, el animal de tipo silvestre del que proviene el animal modificado genéticamente puede comprender un ortólogo correspondiente con la proteína humana funcional. En este caso, la secuencia ortóloga en el animal "humanizado" está inactivada de manera que no se crean proteínas funcionales y el animal "humanizado" comprende al menos una secuencia integrada en el cromosoma 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 el patrón de expresión de la proteína está alterado. Por ejemplo, las regiones reguladoras que controlan la expresión de la proteína, tal como un promotor o un sitio de unión a factor de transcripción, se pueden alterar de manera que la proteína se sobreproduce, o se modifique la expresión temporal o específica de tejido de la proteína, o una combinación de los mismos. Como alternativa, el patrón de expresión de la proteína se puede alterar usando un sistema de animal genomodificado secundariamente por supresión génica. Un ejemplo no limitante de un sistema de animal genomodificado secundariamente por supresión génica incluye un sistema de recombinación Cre-lox. Un sistema de recombinación Cre-lox comprende una enzima recombinasa Cre, una ADN recombinasa 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 expresión temporal y específica de tejido son conocidos en la materia. En general, un animal modificado genéticamente se genera con sitios lox que flanquean una secuencia cromosómica. En animal modificado genéticamente que comprende una secuencia cromosómica flanqueada por lox se puede cruzar entonces con otro animal modificado genéticamente que exprese la recombinasa Cre. Los animales de la prole que comprende la secuencia cromosómica flanqueada por lox y la Cre recombinasa se producen entonces, y la secuencia cromosómica flanqueada por lox se recombina, llevando a una delección o inversión de la secuencia cromosómica que codifica la proteína. La expresión de la recombinasa Cre se puede regular de manera temporal y secundaria para efectuar una recombinación regulada de manera temporal y secundaria de la secuencia cromosómica.

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

Los animales modificados genéticamente desvelados en el presente documento se pueden cruzarse para crear animales que comprendan más de una secuencia cromosómica modificada o para crear animales que sean homocigóticos para una o más secuencias cromosómicas modificadas. Por ejemplo, dos animales que comprenden la misma secuencia cromosómica modificada se pueden cruzar para crear un animal homocigótico para la secuencia cromosómica modificada. Como alternativa, los animales con diferentes secuencias cromosómicas modificadas se pueden cruzar para crear un animal que comprenda ambas secuencias cromosómicas modificadas.

Por ejemplo, un primer animal que comprende un gen "x" de una secuencia cromosómica inactivada se puede cruzar con un segundo animal que comprende una secuencia integrada en el cromosoma que codifica una proteína "X" de gen humano para dar lugar a una descendencia "X" de genes "humanizados" que comprende tanto la secuencia cromosómica del gen inactivado "x" como la secuencia "X" del gen humano integrado en el cromosoma. Asimismo, un animal con gen "X" humanizado se puede cruzar con un animal con gen "Y" humanizado para crear una descendencia con genes X/Y humanizados. Los expertos en la materia apreciarán que son posibles muchas combinaciones.

En otras realizaciones, un animal que comprende una secuencia cromosómica modificada se puede cruzar para combinar la secuencia cromosómica modificada con otros acervos genéticos. A modo de ejemplo no limitante, otros acervos genéticos pueden incluir los acervos genéticos de tipo silvestre, los acervos genéticos con mutaciones de delección, los acervos genéticos con otra integración dirigida y los acervos genéticos con integraciones no dirigidas.

El término "animal", tal 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, moluscos y peces. Los ejemplos de mamíferos adecuados incluyen, sin limitación, roedores, animales de compañía, ganado y primates. Los ejemplos no limitantes de roedores incluyen ratones, ratas, hámsteres, jerbos y cobayas. Los animales de compañía adecuados incluyen, pero sin limitación, 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, pero sin limitación, monos capuchinos, chimpancés, lemures, 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 nemátodo y similares. Los ejemplos no limitantes de insectos incluyen Drosophila y mosquitos. Un animal ejemplar es una rata. Los ejemplos no limitantes de las cepas de rata incluyen Dahl Salt-Sensitive, Fischer 344, Lewis, Long

Evans Hooded, Sprague-Dawley y Wistar. En una realización, el animal no es un ratón modificado genéticamente. En cada una de las repeticiones precedentes de animales adecuados para la invención, el animal no incluye secuencias de transposones introducidas de forma exógena e integradas de manera aleatorizada.

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

Las células son células eucariotas. Las células hospedadoras adecuadas incluyen hongos o levaduras, tales como *Pichia*, *Saccharomyces*, o *Schizosaccharomyces*; células de insecto, tales como células SF9 *Spodoptera frugiperda* o células S2 de *Drosophila melanogaster*; y células de animales, tales como ratón, rata, hámster, primate no humano o células humanas. 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 las roturas de doble cadena. Las células pueden ser de una variedad de tipos celulares, por ejemplo, fibroblastos, mioblastos, linfocitos T o B, macrófagos, células epiteliales, etcétera.

Cuando se usan líneas celulares de mamífero, la línea celular puede ser cualquier línea celular establecida o una línea de células primarias que no se haya descrito aún. La línea celular puede ser adherente o no adherente o la línea celular puede crecer en condiciones que estimulen el crecimiento adherente, no adherente u organotípico usando técnicas convencionales conocidas por los expertos en la materia. Los ejemplos no limitantes de células y líneas celulares de mamífero se proporcionan en el presente documento en la sección (IV)(g). Aún en otras realizaciones, la célula puede ser una célula madre. Los ejemplos no limitantes de células madre adecuadas se proporcionan en la sección (IV)(g).

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 tal como se describe anteriormente en el presente documento (en los párrafos que describen las modificaciones de secuencias cromosómicas en animales) usando los procedimientos descritos en el presente documento. En una realización, el embrión es un embrión no humano de etapa unicelular fecundado de la especie animal de interés. Los embriones ejemplares, incluyendo los embriones unicelulares, incluyen sin limitación, de ratón, de rata, de hámster, de roedor, de conejo, de gato, de perro, de oveja, de cerdo, de vaca, de caballo y embriones de primates.

Definiciones

Salvo que se defina de otra manera, todos los términos técnicos y científicos utilizados en el presente documento tienen el significado comúnmente entendido por un experto en la materia a la cual pertenece la presente invención. Las siguientes referencias a un experto en la materia una definición general de muchos de los términos usados en la presente invención: Singleton y col., Dictionary of Microbiology and Molecular Biology (2ª ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5ª Ed., R. Rieger y col. (eds.), Springer Verlag (1991); y Hale y 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 atribuyen a menos que se especifique lo contrario.

Cuando se introducen elementos de la presente divulgación o de la(s) realización(es) preferida(s) de la(s) misma(s), los artículos "un", "una", "el", "la" y "dicho" pretenden significar que son uno o más de los elementos. Las expresiones "que comprende", "que incluye" y "que tiene" pretenden ser inclusivas y significar que pueden ser elementos adicionales que no sean los elementos enumerados.

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

El término "exógena", tal como se usa en el presente documento, se refiere a una secuencia que se refiere a una secuencia que no es natural para la célula, o una secuencia cuya localización natural en el genoma de la célula está en una localización cromosómica diferente.

Un "gen", tal como se usa en el presente documento, se refiere a una región de ADN (incluyendo exones e intrones) que codifica un producto génico, así como a todas las regiones de ADN que regulan la producción del producto génico, tanto si tales secuencias reguladoras son adyacentes o no a secuencias codificantes y/o transcritas. En consecuencia, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada interna en ribosomas, potenciadores, silenciadores, aislantes, elementos limitantes, orígenes de replicación, sitios de unión a matriz y regiones de control de locus.

El término "heterólogo" se refiere a una entidad que no es endógena o natural para la célula de interés. Por ejemplo,

una proteína heteróloga se refiere a una proteína que proviene de o que originalmente proviene de una fuente exógena, tal como una secuencia de ácido nucleico introducida de manera exógena. En algunos casos, la proteína heteróloga no se produce normalmente por la célula de interés.

Las expresiones "ácido nucleico" y "polinucleótido" se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, de 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. Estos términos pueden abarcar los análogos conocidos de los nucleótidos naturales, así como nucleótidos que se modifican en la base, restos de azúcar y/o fosfatos (por ejemplo, estructuras principales de fosforioatos). En general, un análogo de un nucleótido particular tiene la misma especificidad de emparejamiento de bases; es decir, un análogo de A emparejará su base con T.

El término "nucleótido" se refiere a desoxirribonucleótidos o ribonucleótidos. Los nucleótidos pueden ser nucleótidos convencionales (es decir, adenosina, guanosina, citidina, timidina y uridina) o análogos de nucleótido. Un análogo de nucleótido se refiere a un nucleótido que tiene una base de purina o de pirimidina modificada o un resto de 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 de azúcar o de base de un nucleótido incluyen la adición (o eliminación) de grupos acetilo, grupos amino, grupos carboxilo, grupos carboximetilo, grupos hidroxilo, grupos metilo, grupos fosforilo y grupos tiol, así como la sustitución de los átomos de carbono y nitrógeno de las bases por otros átomos (por ejemplo, 7-deaza purinas). Los análogos de nucleótidos también incluyen didesoxinucleótidos, 2'-O-metil nucleótidos, ácidos nucleicos bloqueados (LNA), ácidos peptidonucleicos (PNA) y morfolinos.

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

Las técnicas para determinar la identidad de secuencia de ácido nucleico y de aminoácidos se conocen en la materia. Normalmente, tales técnicas incluyen la determinación de la secuencia de nucleótidos de un ARNm para un gen y/o la determinación de la secuencia de aminoácidos codificada por la misma, y la comparación de estas secuencias con una segunda secuencia de nucleótidos o de aminoácidos. Las secuencias genómicas también se pueden determinar y comparar de esta manera. En general, identidad se refiere a una correspondencia exacta nucleótido a nucleótido o aminoácido a aminoácido de dos polinucleótidos o secuencias polipeptídicas, respectivamente. Dos o más secuencias (de polinucleótidos o aminoácidos) se pueden comparar determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, sean secuencias de ácido nucleico o de aminoácidos, es el número de coincidencias exactas entre dos secuencias alineadas dividido entre la longitud de las secuencias más cortas y multiplicado por 100. Un alineamiento aproximado para las secuencias de ácido nucleico se proporciona mediante el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). Este algoritmo se puede aplicar a las secuencias de aminoácidos usando la matriz de puntuación desarrollada por Dayhoff, *Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 supl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., EEUU, y normalizada por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). Una implementación ejemplar de este algoritmo para determinar el porcentaje de identidad de una secuencia se proporciona por el Genetics Computer Group (Madison, Wis.) en la aplicación de utilidad "BestFit". Otros programas adecuados para calcular el porcentaje de identidad o la similitud entre secuencias son generalmente conocidos en la materia, por ejemplo, otro programa de alineamiento es BLAST, usado con parámetros por defecto. Por ejemplo, se pueden usar BLASTN y BLASTP usando los siguientes parámetros por defecto: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Los detalles de estos programas se pueden encontrar en la página web de GenBank.

Como se pueden hacer diversos cambios en las células mencionadas y procedimientos sin apartarse del ámbito de la invención, se pretende que toda la materia contenida en la descripción anterior y en los ejemplos que figuran a continuación, se puedan interpretar como ilustrativos y no en un sentido limitante.

Ejemplos ilustrativos

Los siguientes ejemplos ilustran determinados aspectos de la divulgación.

Ejemplo ilustrativo 1: Modificación del gen Cas9 para la expresión en mamíferos

Un gen Cas9 de la cepa MGAS15252 de *Streptococcus pyogenes* (Número de referencia YP_005388840.1) se optimizó con el codón de *Homo sapiens* de preferencia para potenciar su traducción en células de mamífero. El gen Cas9 también se modificó añadiendo una señal de localización PKKKRKV (SEQ ID NO:1) en el extremo C-terminal para dirigir la proteína a los núcleos de las células de mamífero. La Tabla 1 presenta 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 y optimizada con codón.

Tabla 1. Secuencia de aminoácidos de Cas9 modificada

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGET
 AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHER
 HPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
 PDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRENLIAQLPGEKR
 NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLA
 AKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK
 SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTK
 VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
 DRFNASLGAYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFD
 DKVMKQLKRRRYTGWGRLSRKLINGIRDQSGKTILDFLKSDGFANRNFQMQLIHDDS
 LTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVI
 EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
 GRDMYVDQELDINRLSDYVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEV
 VKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
 QILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLN
 AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKT
 EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFS
 KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLKSVKEL
 LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK
 GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
 LADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQSI TGLYETRIDLSQLGGDPKKKRKV (SEQ ID NO:9)

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

ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTG
 GGCCGTGATCACCGACGACTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGG
 GCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTC
 GGCTCTGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAA
 GATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACG
 AGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGG
 TGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGAC
 GAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAAGCTG
 GCCGACAGCACCGACAAGGCCGACCTGAGACTGATCTACCTGGCCCTGGCCCA
 CATGATCAAGTTCGGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAA
 CAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGATCTACAATCAGCTGTT
 CGAGGAAAACCCCATCAACGCCAGCAGAGTGGACGCCAAGGCCATCCTGAGCG
 CCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCCGGC
 GAGAAGCGGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGAC
 CCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAG
 CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACC

(continuación)

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

AGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGA
 GCGACATCCTGAGAGTGAACAGCGAGATACCAAGGCCCCCCTGTCCGCCTCT
 ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTC
 GTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAG
 AACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA
 GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCAGGAAGTCTCGTGAA
 GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCA
 TCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAA
 GATTTTTACCCATTCTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCT
 TCAGAATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCTGCCT
 GGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTG
 GTGGACAAGGGCGCCAGCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGA
 TAAGAACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTA
 CTTACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGCG
 GAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGACCTGCTGT
 TCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGA
 AAATCGAGTGCTTCGACAGCGTGGAATCAGCGGCGTGGAAGATCGGTTCAACG
 CCTCCCTGGGCGCCTATCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCC
 TGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACAC
 TGTTTGAGGACCGGGGCATGATCGAGGAACGGCTGAAAACCTATGCCACCTGT
 TCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
 AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGAC
 AATCCTGGATTTCTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCT
 GATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTC
 CGGCCAGGGACACTCTCTGCACGAGCAGATCGCCAATCTGGCCGGATCCCCCG
 CCATTAAGAAGGGCATCCTGCAGACAGTGAAGATTGTGGACGAGCTCGTGAAAG
 TGATGGGCCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
 ACCACCCAGAAGGGACAGAGAAGAACAGCCGCGAGAGAAATGAAGCGGATCGAAGA
 GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
 CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAAATGGGCGGGATA
 TGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGACTACGATGTGGACC
 ACATTGTGCCCCAGTCCTTCATCAAGGACGACTCCATCGATAACAAAGTGCTGAC
 TCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTC
 GTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACC
 CAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGGCGGCCTGAGCGAACT
 GGATAAGGCCGGCTTCATTAAGCGGCAGCTGGTGGAACCCGGCAGATCACAA
 AGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAC
 GACAACTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCC
 GACTTCAGAAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACC
 ACGCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
 TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGATTACAAGGTGTACGACGTG
 CGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTA
 CTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATCACACTGGCCAAC
 GGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGAT
 CGTGTGGGATAAGGGCCGGGATTTTGCCACAGTGCGGAAAGTGCTGTCCATGC
 CCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACCGGCGGCTTCAGCAAA

(continuación)

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

GAGTCTATCCTGCCCAAGAGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGAT
 TGGGACCCTAAGAAGTACGGCGGCTTTGACAGCCCCACCGTGGCCTACTCTGT
 GCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGA
 AAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCA
 TCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCA
 AGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGCGGATGCTG
 GCTTCTGCCGGCGAACTGCAGAAGGGAAACGAGCTGGCCCTGCCCTCCAAATA
 TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGA
 GGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGA
 GATCATCGAGCAGATTAGCGAGTTCTCCAAGCGCGTGATCCTGGCCGATGCCAA
 CCTGGACAAGGTGCTGAGCGCCTACAACAAGCACCGGGATAAGCCCATCAGAG
 AGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTG
 CCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCA
 AAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
 ACACGGATCGACCTGTCTCAGCTGGGAGGCGACCCCAAGAAAAAGCGCAAAGT
 G (SEQ ID NO:10)

La secuencia de ADN de Cas9 modificada se colocó en 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 en el promotor de control T7 para la síntesis *in vitro* de ARNm con la ARN polimerasa T7. La transcripción *in vitro* de ARN se realizó usando el kit MessageMAX T7 ARCA-Capped Message Transcription y el sistema de producción de ARNm T7 mScript Standard (Cellscript).

Ejemplo ilustrativo 2: Cas9 de direccionamiento

El locus del sitio de integración 1 del virus adenoasociado (AAVS1) se usó como una diana para la modificación del genoma humano mediada por Cas9. El locus AAVS1 humano se localiza en el intrón 1 (4427 pb) de la proteína fosfatasa 1, subunidad reguladora 12C (PPP1 R12C). La Tabla 3 presenta el primer exón (sombreado en gris) y el primer intrón de PPP1 R12C. La secuencia subrayada en el intrón es el sitio de modificación dirigida (es decir, el locus AAVS1).

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

GCGGGCGGGCGGTGCGATGTCCGGAGAGGATGGCCCGGCGGCTGGCCCGGG
 GGCGGGCGGGCGGGCGGCTGCCCGGGAGCGGCGACGGGAGCAGCTGCGGCAGTG
 GGGGGCGCGGGCGGGCGCCGAGCCTGGCCCCGGAGAGCGCCGCGCCCGCAC
 CGTCCGCTTCGAGCGCGCCGCGGAGTTCTTGGCGGCCTGTGCGGGCGGGCGAC
 CTGGACGAGGCGCGTCTGATGCTGCGCGCCGCGGACCTGGCCCGGGCGCCG

(continuación)

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

AGCTCGACCCCGCCGCGCCGCCGCCGCCGCCGCGCCGCTGCTGGACTCCACCAA
CGCCGACGGTATCAGCGCCCTGCACCAAGGTACGCGCCCCCGCCCGGCGTCT
CCCGGGGCCAGGTCCACCCTCTGCTGCGCCACCTGGGGCATCCTCCTTCCCCG
TTGCCAGTCTCGATCCGCCCCGTGTTCTTGGCCCTGGGCTTTGCCACCCTATG
CTGACACCCCGTCCCAGTCCCCCTTACCATTCCCCCTTCGACCACCCCACTTCCG
AATTGGAGCCGCTTCAACTGGCCCTGGGCTTAGCCACTCTGTGCTGACCACTCT
GCCCCAGGCCTCCTTACCATTCCCCTTCGACCTACTCTCTTCCGCATTGGAGTC
GCTTTAACTGGCCCTGGCTTTGGCAGCCTGTGCTGACCCATGCAGTCCCTCTTA
CCATCCCTCCCTCGACTTCCCCCTCTTCCGATGTTGAGCCCCCTCCAGCCGGTCCT
GGACTTTGTCTCCTTCCCTGCCCTGCCCTCTCCTGAACCTGAGCCAGCTCCCAT
AGCTCAGTCTGGTCTATCTGCCTGGCCCTGGCCATTGTCACTTTGCGCTGCCCT
CCTCTCGCCCCCGAGTGCCCTTGCTGTGCCGCCGGAACCTCTGCCCTCTAACGCT
GCCGTCTCTCTCCTGAGTCCGGACCACTTTGAGCTCTACTGGCTTCTGCGCCGC
CTCTGGCCCACTGTTTCCCCTTCCCAGGCAGGTCTGTCTTCTCTGACCTGCATT
CTCTCCCCTGGGCCTGTGCCGCTTTCTGTCTGCAGCTTGTGGCCTGGGTACCT
CTACGGCTGGCCAGATCCTTCCCTGCCGCCTCCTTCAGGTTCCGTCTTCTCTCC
ACTCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGCCCAAGGATGCTCTTTCCGG
AGCACTTCTTCTCGGCGCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCC
GTGTCTGGGTCTCTCCGGGCATCTCTCCTCCCTCACCCAACCCCATGCCGTCT
TCACTCGCTGGGTTCCCTTTTCTTCTCCTTCTGGGGCCTGTGCCATCTCTCGTT
TCTTAGGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTCCCGCCTCCCCTTCT
TGTAGGCCTGCATCATCACCGTTTTTCTGGACAACCCCAAAGTACCCCGTCTCCC
TGGCTTTAGCCACCTCTCCATCCTCTTGCTTTCTTGGCCTGGACACCCCGTTCTC
CTGTGGATTGGGTACCTCTCACTCCTTTCATTGGGCAGCTCCCCTACCCCC
CTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGTCATGGCATCTTCCAG
GGGTCCGAGAGCTCAGCTAGTCTTCTTCTCCAACCCGGGCCCTATGTCCACT
TCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCA
CCTTATATTCCCAGGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGT
CCCCCTCCACCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCC
CCATCCTTAGGCCTCCTCCTTCCCTAGTCTCCTGATATTGGGTCTAACCCCCACCT
CCTGTTAGGCAGATTCTTATCTGGTGACACACCCCCATTTCCTGGAGCCATCTC
TCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGGATCCTG
GGAGGGAGAGCTTGGCAGGGGGTGGGAGGGAAGGGGGGGATGCGTGACCTG
CCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCAGAACCTGAGCTGCTCT
GACGCGGCCGTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCTTACACT
TCCCAAGAGGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAGTTGGTCCTGAG
TTCTAACTTTGGCTCTTACCTTTCTAGTCCCCAATTTATATTGTTCTCCGTGCG
TCAGTTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCTGGCAGGGCTGT
GGTGAGGAGGGGGGTGTCCGTGTGGAAACTCCCTTTGTGAGAATGGTGCGTC
CTAGGTGTTACCAAGGTCGTGGCCGCCTCTACTCCCTTTCTCTTTCTCCATCCTT
CTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACATATTCTCCGCCAGAGCA
GGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGAGTCCTTGGCA
AGCCCAGGAGAGGCGCTCAGGCTTCCCTGTCCCCCTTCTCGTCCACCATCTCA
TGCCCCCTGGCTCTCCTGCCCTTCCCTACAGGGGTTCTGGCTCTGCTCTTCAG
ACTGAGCCCCGTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTTCCCCTGCAT
CCCCCAGAGGGCCCCAGGCCACCTACTTGGCCTGGACCCCCACGAGAGGGCCACCC

(continuación)

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

CAGCCCTGTCTACCAGGCTGCCTTTTGGGTGGATTCTCCTCCAACGTGTGGGGTG
 ACTGCTTGGCAAACCTCACTCTTCGGGGTATCCCAGGAGGCCTGGAGCATTGGG
 GTGGGCTGGGGTTTCAGAGAGGAGGGATTCCCTTCTCAGGTTACGTGGCCAAGA
 AGCAGGGGAGCTGGGTTTGGGTGAGGTCTGGGTGTGGGGTGACCAGCTTATGC
 TGTTCGCCAGGACAGCCTAGTTTTAGCACTGAAACCCTCAGTCCTAGGAAAACA
 GGGATGGTTGGTCACTGTCTCTGGGTGACTCTTGATTCCCGGCCAGTTTCTCCA
 CCTGGGGCTGTGTTTCTCGTCCTGCATCCTTCTCCAGGCAGGTCCCCAAGCATC
 GCCCCCTGCTGTGGCTGTTCCCAAGTTCTTAGGGTACCCACGTGGGTTTATC
 AACCCTTGGTGAGGCTGGTACCCTGCCCCATTCTGCACCCCAATTGCCTTA
 GTGGCTAGGGGGTTGGGGGCTAGAGTAGGAGGGGCTGGAGCCAGGATTCTTAG
 GGCTGAACAGAGAAGAGCTGGGGGCCTGGGCTCCTGGGTTTGAGAGAGGAGG
 GGCTGGGGCCTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCCTGGACT
 CCTGGGTCTGAGGGTGGAGGGACTGGGGGCCTGGACTCCTGGGTCCGAGGGA
 GGAGGGGCTGGGGCCTGGACTCGTGGGTCTGAGGGAGGAGGGGCTGGGGGC
 CTGGACTTCTGGGTCTTAGGGAGGCGGGGCTGGGCCTGGACCCCTGGGTCTGA
 ATGGGGAGAGGCTGGGGGCCTGGACTCCTTCATCTGAGGGCGGAAGGGCTGG
 GGCCTGGCCTCCTGGGTTGAATGGGGAGGGGTTGGGCCTGGACTCTGGAGTCC
 CTGGTGCCAGGCCTCAGGCATCTTTCACAGGGATGCCTGTACTGGGCAGGTC
 CTTGAAAGGGAAAGGCCATTGCTCTCCTTGCCCCCTCCCCTATCGCCATGAC
 AACTGGGTGGAAATAAACGAGCCGAGTTCATCCCGTTCCAGGGCACGTGCGG
 CCCCTTCACAGCCCGAGTTTCCATGACCTCATGCTCTTGGCCCTCGTAGCTCCC
 TCCCGCCTCCTCCAGATGGGCAGCTTTGGAGAGGTGAGGGACTTGGGGGGTAA
 TTTATCCCGTGGATCTAGGAGTTTAGCTTCACTCCTTCCTCAGCTCCAGTTCAGG
 TCCCGGAGCCCACCCAGTGTCCACAAGGCCTGGGGCAAGTCCCTCCTCCGACC
 CCCTGGACTTCGGCTTTTGTCCCCCAAGTTTGGACCCCTAAGGGAAGAATGA
 GAAACGGTGGCCCGTGTGACCCCTGGCTGCAGGGCCCCGTGCAGAGGGGGC
 CTCAGTGAACCTGGAGTGTGACAGCCTGGGGCCCAGGCACACAGGTGTGCAGCT
 GTCTACCCCTCTGGGAGTCCCGCCCAGGCCCTGAGTCTGTCCCAGCACAGG
 GTGGCCTTCCTCCACCCTGCATAGCCCTGGGCCACGGCTTCGTTCTGCAGA
 GTATCTGCTGGGGTGGTTTCCGAGCTTGACCCTTGGAAGGACCTGGCTGGGTTT
 AAGGCAGGAGGGGCTGGGGGCCAGGACTCCTGGCTCTGAAGGAGGAGGGGCT
 GGAACCTCTTCCCTAGTCTGAGCACTGGAAGCGCCACCTGTGGGTGGTGACGG
 GGGTTTTGCCGTGTCTAACAGGTACCATGTGGGGTTCCCGCACCCAGATGAGAA
 GCCCCCTCCCTTCCCCGTTCACTTCCTGTTTGCAGATAGCCAGGAGTCCTTTCGT
 GGTTTCCACTGAGCACTGAAGGCCTGGCCGGCCTGACCACTGGGCAACCAGGC
 GTATCTTAACAGCCAGTGGCCAGAGGCTGTTGGGTCAATTTCCCACTGTCTTA
 GCACCGTGTCCCTGGATCTGTTTTCTGGGTCCCTCTGGAGTCCCGACTTGCTG
 GGACACCGTGGCTGGGGTAGGTGCGGCTGACGGCTGTTTCCCACCCCCAG
 (SEQ ID NO:11)

Los ARN guías de Cas9 se diseñaron para el direccionamiento del locus humano AAVS1. Un ARN de 42 nucleótidos (denominado en el presente documento como una secuencia "ARNcr") que comprende (de 5' a 3') una secuencia de reconocimiento diana (es decir, una secuencia complementaria con la hebra no codificante de la secuencia diana) y secuencia de protoespaciador; un ARN de 85 nucleótidos (denominado en el presente documento como una secuencia de "ARNtracr") que comprende la secuencia 5' con complementariedad con la secuencia 3' del ARNcr y una secuencia de horquilla adicional; y un ARN quimérico que comprende los nucleótidos 1-32 del ARNcr, un bucle GAAA y los nucleótidos 19-45 del ARNtracr se preparó. El ARNcr se sintetizó químicamente por Sigma-Aldrich. El ARNtracr y el ARN quimérico se sintetizó mediante transcripción *in vitro* con ARN polimerasa T7 usando el kit T7-Scribe Standard RNA IVT (Cellscript). La secuencia codificante de ARN quimérico también se colocó en el control del promotor humano U6 para la transcripción *in vivo* en células humanas. La Tabla 4 presenta las secuencias de los ARN guías.

Tabla 4. ARN guías		
ARN	Secuencia 5'-3'	SEQ ID NO:
AAVS1-ARNcr	ACCCACAGUGGGGCCACUAGUUUUAGAGCUAUGCUGU UUUG	12
ARNtracr	GGAACCAUUCAAAACAGCAUAGCAAGUUAUUUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUU	13
ARN quimérico	ACCCACAGUGGGGCCACUAGUUUUAGAGCUAGAAUA GCAAGUUAUUUAAGGCUAGUCCG	14

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

Se usó la integración dirigida de una proteína GFP en el extremo N-terminal de PPP1 R12C para controlar la modificación del genoma mediada por Cas9. Para mediar la integración mediante recombinación homóloga se preparó un polinucleótido donador. El donador de ADN de AAVS1-GFP contenía en 5' un brazo homólogo del locus AAVS1 (1185 pb), un receptor de corte y empalme de ARN, una secuencia codificante de turbo GFP, un finalizador de transcripción en 3', y un brazo homólogo del locus AAVS1 en 3' (1217 pb). La Tabla 5 presenta las secuencias del receptor de corte y empalme de ARN y la secuencia codificante de GFP seguida por el finalizador de transcripción en 3'. El ADN del plásmido se preparó usando el kit GenElute Endotoxin-Free Plasmid Maxiprep (Sigma).

Tabla 5. Secuencias en la secuencia del donador de ADN AAVS1.GFP		
	Secuencia 5'-3'	SEQ ID NO:
Receptor de corte y empalme de ARN	CTGACCTCTTCTCTCCTCCACAG	15
Secuencia codificante de GFP y finalizador de transcripción	GCCACCATGGACTACAAAGACGATGACGACAAGGTCTGACT CTAGAGCTGCAGAGAGCGACGAGAGCGGCCTGCCCCGCCA TGGAGATCGAGTGCCGCATCACCGGCACCCTGAACGGCG TGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGCACCCCCG AGCAGGGCCGCATGACCAACAAGATGAAGAGCACCAAAGG CGCCCTGACCTTCAGCCCCTACCTGCTGAGCCACGTGATG GGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCT ACGAGAACCCCTTCCTGCACGCCATCAACAACGGCGGCTA CACCAACACCCGCATCGAGAAGTACGAGGACGGCGGCGT GCTGCACGTGAGCTTCAGCTACCGCTACGAGGCCGGCCG CGTGATCGGCGACTTCAAGGTGATGGGCACCGGCTTCCCC GAGGACAGCGTGATCTTCACCGACAAGATCGTCCGCAGCA ACGCCACCGTGGAGCACCTGCACCCCATGGGCGATAACG ATCTGGATGGCAGCTTCACCCGCACCTTCAGCCTGCGCGA CGGCGGCTACTACAGCTCCGTGGTGGACAGCCACATGCAC TTCAAGAGCGCCATCCACCCCAGCATCTGCAGAACGGGG GCCCCATGTTGCCTTCCGCCGCGTGGAGGAGGATCACA GCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCCTT CAAGACCCCGGATGCAGATGCCGGTGAAGAATGAAGATCT CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTCCCCCTCC CCCGTGCCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA GTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGG ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG CTGGGGATGCGGTGGGCTCTATGGAAGTTCGAGGTTAAACG TCGACGCGGCCGCGT	16

La integración de genes dirigida dará como resultado una proteína de fusión entre los primeros 107 aminoácidos de la PPP1 R12C y la turbo GFP. La proteína de fusión esperada contiene los primeros 107 restos de aminoácidos de

PPP1R12C (resaltado en gris) de corte y empalme de ARN entre el primer exón de PPP1 R12C y el receptor de corte y empalme diseñado por ingeniería genética (véase la Tabla 6).

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

MSGEDGPAAGPGAAAAAARERRREQLRQWGARAGAEPGPGERRARTVRFERAAE
FLAACAGGDLDEARLMLRAADPGPGAELDPAAPPPARAVLDSTNADGISALHQATM
DYKDDDDKVDSRAAESDESGLPAMEIECRITGTLNGVEFELVGGGEGTPEQGRMTN
KMKSTKGALTFSPYLLSHVMGYGFYHFGTYPSTGYENPFLHAINNGGYTNTRIEKYED
GGVLHVSFSYRYEAGRVIGDFKVMGTGFPEDSVIFTDKIVRSNATVEHLHPMGDNDL
DGSFTRTFSLRDGGYYSSVVDSHMHFKSAIHPSILQNGGPMFAFRRVEEDHSNTEL
GIVEYQHAFKTPDADAGEE (SEQ ID NO:17)

Ejemplo ilustrativo 4: Integración dirigida mediada por Cas9

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 FBS al 10 % 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 (a aproximadamente 0,5 millones de células por ml antes de la transfección). Las células se transfectaron con Solución V de Nucleofector (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 CO₂ al 5 % inmediatamente tras la nucleofección.

Tabla 7. Tratamientos de transfección.

Tratamiento	Cas9 modificada	ARN guía	Secuencia donadora
A	ARNm de Cas9 transcrito con un análogo de casquete anti-inverso (10 µg)	doble cadena de ARNcr-ARNtracr prealineada (0,3 nmol)	ADN del plásmido AAVS1-GFP (10 µg)
B	ARNm de Cas9 transcrito con un análogo de casquete anti-inverso (10 µg)	ARN quimérico (0,3 nmol)	ADN del plásmido AAVS1-GFP (10 µg)
C	ARNm de Cas9 con casquete mediante la reacción de postrancripción de casquete (10 µg)	ARN quimérico (0,3 nmol)	ADN del plásmido AAVS1-GFP (10 µg)
D	ADN del plásmido Cas9 (10 µg)	ADN del plásmido de ARN quimérico con U6 (5 µg)	ADN del plásmido AAVS1-GFP (10 µg)
E	Ninguno	Ninguno	ADN del plásmido AAVS1-GFP (10 µg)
F	Ninguno	Ninguno	Ninguno

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

Ejemplo ilustrativo 5: Confirmación por PCR de la integración dirigida

El ADN genómico se extrajo de las células transfectadas con el kit GenElute Mammalian Genomic DNA Miniprep (Sigma) 12 días después de la transfección. El ADN genómico se amplificó entonces por PCR con un cebador hacia delante localizado fuera del brazo homólogo en 5' del donador del plásmido AAVS1-GFP y un cebador inverso localizado en la región 5' de la GFP. El cebador hacia delante fue 5'-CCACTCTGTGCTGACCACTCT-3' (SEQ ID NO:18) y el cebador inverso fue 5'-GCGGCACTCGATCTCCA-3' (SEQ ID NO:19). El tamaño del fragmento esperado de la PCR de unión fue de 1388 pb. La amplificación se llevó a cabo con JumpStart Taq ReadyMix (Sigma), usando las siguientes condiciones de ciclación: 98 °C durante 2 minutos para la desnaturalización 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 %.

Las células transfectadas con 10 µg de ARNm de Cas9 transcrito con un análogo de casquete anti-inverso, 0,3 nmol de doble cadena de ARNcr-ARNtracr prealineada, y 10 µg ADN del plásmido AAVS1-GFP presentaron un producto de PCR del tamaño esperado (véase el carril A, FIG. 3).

Ejemplo ilustrativo 6: Modificación de genoma en embriones de ratón basada en Cas9

El locus Rosa26 de ratón se puede dirigir para modificaciones genómicas. La Tabla 8 presenta una porción de la secuencia de Rosa26 de ratón en la que los posibles sitios diana se muestran en negrita. Cada sitio diana comprende un protoespaciador.

Tabla 8. Secuencia Rosa26 de ratón	
GAGCGGCTGCGGGGCGGGTGCAAGCACGTTTCCGACTTGAGTTGCCTCAAGAG GGGCGTGCTGAGCCAGACCTCCATCGCGCACTCCGGGGAGTGGAGGGAAGGA GCGAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCCAAAGT	
CGCTCTGAGTTGTTATCAGTAAGGGAGCTGCAGTGGAGTAGGCGGGGAGAAGG CCGCACCCTTCTCCGGAGGGGGGAGGGGAGTGTGCAATACCTTTCTGGGAGT TCTCTGCTGCCTCCTGGCTTCTGAGGACCGCCCTGGGCCTGGGAGAATCCCTTC CCCCTCTTCCCTCGTGATCTGCA ACTCCAGTCTTTCTAGAAGAT GGGCGGGAGT CTTCTGGGCAGGCTTAAAGGCTAACCTGGTGTGTGGGCGTTGTCCTGCAGGGG AATT GAACAGGTGTA AAATT GG AGGGACAAGACTTCCACAGATTTTCGGTTTT GT CGGGAAGTTTTT TA ATAGGGGCAAATAAGGAAAATGGGAGGATAGGTAGTCA TCTGGGGTTTTATGCAGCAAACTACAGGTTATTATTGCTTGTGATCCGCCTCGG AGTATTTTCCATCGAGGTAGATTAAAGACATGCTCACCCGAGTTTTATACTCTCCT GCTTGAGATCCTTACTACAGTATGAAATTACAGTGTGCGGAGTTAGACTATGTAA GCAGAATTTTA (SEQ ID NO:20)	

- 5 Los ARN guías se diseñaron para direccionar cada uno de los sitios diana en el locus Rosa26 de ratón. Estas secuencias se muestran en la Tabla 9, cada una es de 42 nucleótidos de longitud y la región 5' es complementaria a la hebra que no se presenta en la Tabla 8 (es decir, la hebra que es complementaria a la hebra mostrada en la Tabla 8).

Tabla 9. ARN guías 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

- 10 Los ARNcr se sintetizaron químicamente y se prealinearon con el ARNtracr (SEQ ID NO:13; véase el Ejemplo 2). El ARNcr / ARNtracr prealineado y el ARNm transcrito *in vitro* que codifica la proteína Cas9 modificada (SEQ ID NO. 9; véase el Ejemplo 1) se pueden microinyectar en los pronúcleos de embriones de ratón fecundados. Tras la guía hacia la diana establecida por el ARNcr, la proteína Cas9 escinde el sitio diana, y la rotura de doble cadena resultante se puede reparar mediante un proceso de reparación por unión de extremos no homólogos (NHEJ). Los embriones inyectados se pueden incubar bien a 37 °C, con CO₂ al 5 % durante la noche o durante hasta 4 días, seguido por el análisis de genotipado, o los embriones inyectados se pueden implantar en los ratones hembra receptores de manera que se puedan genotipar los animales nacidos vivos. Los embriones incubados *in vitro* o los tejidos de los animales nacidos vivos se pueden explorar para la presencia de una mutación inducida por Cas9 en el locus Rosa usando los procedimientos convencionales. Por ejemplo, los embriones o tejidos de fetos o animales nacidos vivos se pueden recolectar para la extracción y el análisis de ADN. El ADN se puede aislar usando procedimientos convencionales. La región dirigida del locus Rosa26 se puede amplificar por PCR usando cebadores apropiados. Dado que la NHEJ es propensa a errores, las delecciones de al menos un nucleótido, las inserciones de al menos un nucleótido, las sustituciones de al menos un nucleótido o las combinaciones de las mismas pueden tener lugar durante la reparación de la rotura. Las mutaciones se pueden detectar usando procedimientos de genotipado basados en PCR, tales como los ensayos de emparejamiento incorrecto Cel-I y de secuenciación de ADN.

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

El locus Rosa26 se puede modificar en embriones de ratón mediante coinyección de un polinucleótido donador, tal como se detalla anteriormente en la sección (III)(d), junto con el ARNcr / ARNtracr prealineado y el ARNm que codifica la Cas9 modificada tal como se describe anteriormente en el Ejemplo 6. Los embriones incubados *in vitro* o los tejidos de animales nacidos vivos (tal como se describe en el Ejemplo 6) se pueden explorar para un locus Rosa26 usando procedimientos de genotipado basados en PCR, tales como los ensayos RFLP, PCR de unión, y secuenciación de ADN.

Ejemplo ilustrativo 8: Modificación de genoma en embriones de rata basada en Cas9

El locus Rosa26 de rata se puede dirigir para modificaciones genómicas. La Tabla 10 presenta una porción de la secuencia de rata en la que los posibles sitios diana se muestran en negrita. Cada sitio diana comprende un protoespaciador.

Tabla 10. Secuencia Rosa26 de rata

```

GGGATTCCTCCTTGAGTTGTGGCACTGAGGAACGTGCTGAACAAGACCTACATT
GCACTCCAGGGAGTGGATGAAGGAGTTGGGGCTCAGTCGGGTTGTATTGGAGA
CAAGAAGCACTTGCTCTCCAAAAGTCGGTTTGAGTTATCATTAAAGGGAGCTGCAG
TGGAGTAGGCGGAGAAAAGGCCGCACCCTTCTCAGGACGGGGGAGGGGAGTG
TTGCAATACCTTTCTGGGAGTTCTCTGCTGCCTCCTGTCTTCTGAGGACCGCCCT
GGGCCTGGAAGATTCCCTTCCCCCTTCTTCCCTCGTGATCTGCAACTGGAGTCT
TTCTGGAAGATAGGCGGGAGTCTTCTGGGCAGGCTTAAAGGCTAACCTGGTGC
GTGGGGCGTTGTCTGCGAGAGGAATTGAACAGGTGTAAAATTGGAGGGGCAAG
ACTTCCCACAGATTTTCGATTGTGTTGTTAAGTATTGTAATAGGGGCAAATAAGG
GAAATAGACTAGGCACTCACCTGGGGTTTTATGCAGCAAACTACAGGTTATTAT
TGCTTGTGATCCGCCCTGGAGAATTTTTCACCGAGGTAGATTGAAGACATGCCC

ACCCAAATTTTAATATTCTTCCACTTGCGATCCTTGCTACAGTATGAAA (SEQ ID
NO:24)

```

Los ARN guías se diseñaron para direccionar cada uno de los sitios diana en el locus Rosa26 de rata. Estas secuencias se muestran en la Tabla 11, cada una es de 42 nucleótidos de longitud y la región 5' es complementaria a la hebra que no se presenta en la Tabla 10 (es decir, la hebra que es complementaria a la hebra mostrada en la Tabla 10).

Tabla 11. ARN guías de Rosa26 de rata

ARN	Secuencia 5'-3'	SEQ ID NO:
rRosa26-ARNcr-1	AGGGGGAAGGGAAUCUCCAGUUUUAGAGCUA UGCUGUUUUG	25
rRosa26-ARNcr-2	UCUGCAACUGGAGUCUUUCUGUUUUAGAGCUA UGCUGUUUUG	26
rRosa26-ARNcr-3	AGGCGGGAGUCUUCUGGGCAGUUUUAGAGCUA UGCUGUUUUG	27

Los ARNcr se sintetizaron químicamente y se prealinearon con el ARNtracr (SEQ ID NO:13; véase el Ejemplo 2). El ARNcr / ARNtracr prealineado y el ARNm transcrito *in vitro* que codifica la proteína Cas9 modificada (SEQ ID NO. 9; véase el Ejemplo 1) se pueden microinyectar en los pronúcleos de embriones de rata fecundados. Tras la guía hacia el sitio diana por el ARNcr, la proteína Cas9 escinde el sitio diana, y la rotura de doble cadena resultante se puede reparar mediante un proceso de reparación por unión de extremos no homólogos (NHEJ). Los embriones inyectados se pueden incubar bien a 37 °C, con CO₂ al 5 % durante la noche o durante hasta 4 días, seguido por el análisis de genotipado, o los embriones inyectados se pueden implantar en los ratones hembra receptores de manera que se puedan genotipar los animales nacidos vivos. Los embriones incubados *in vitro* o los tejidos de los animales nacidos vivos se pueden explorar para la presencia de una mutación inducida por Cas9 en el locus Rosa usando los procedimientos convencionales. Por ejemplo, los embriones o tejidos de fetos o animales nacidos vivos se pueden recolectar para la extracción y el análisis de ADN. El ADN se puede aislar usando procedimientos convencionales. La región dirigida del locus Rosa26 se puede amplificar por PCR usando cebadores apropiados. Dado que la NHEJ

es propensa a errores, las deleciones de al menos un nucleótido, las inserciones de al menos un nucleótido, las sustituciones de al menos un nucleótido o las combinaciones de las mismas pueden tener lugar durante la reparación de la rotura. Las mutaciones se pueden detectar usando procedimientos de genotipado basados en PCR, tales como los ensayos de emparejamiento incorrecto Cel-I y de secuenciación de ADN.

5 **Ejemplo ilustrativo 9: Modificación de genoma en embriones de rata basada en Cas9**

El locus Rosa26 se puede modificar en embriones de rata mediante coinyección de un polinucleótido donador, tal como se detalla anteriormente en la sección (III(d)), junto con el ARNcr / ARNtracr prealineado y el ARNm que codifica la Cas9 modificada tal como se describe anteriormente en el Ejemplo 8. Los embriones incubados *in vitro* o los tejidos de ratas nacidas vivas (tal como se describe en el Ejemplo 8) se pueden explorar para un locus Rosa26 usando procedimientos de genotipado basados en PCR, tales como los ensayos RFLP, PCR de unión, y secuenciación de ADN.

LISTADO DE SECUENCIAS

15 <110> SIGMA-ALDRICH CO. LLC CHEN, Fuqiang DAVIS, Gregory D. KANG, Qiaohua KNIGHT, Scott W.

<120> MODIFICACIÓN Y REGULACIÓN DEL GENOMA EN BASE A CRISPR

<130> P2502EP06

20 <140> PCT/US2013/073307

<141> 05-12-2013

<150> US 61/734,256

<151> 06-12-2012

25 <150> US 61/758,624

<151> 30-01-2013

30 <150> US 61/761,046

<151> 05-02-2013

<150> US 61/794,422

<151> 15-03-2013

35 <160> 27

<170> PatentIn versión 3.5

<210> 1

40 <211> 7

<212> PRT

<213> Secuencia artificial

<220>

45 <223> SINTETIZADO

<400> 1

Pro Lys Lys Lys Arg Lys Val
1 5

50 <210> 2

<211> 7

<212> PRT

<213> Secuencia artificial

55 <220>

<223> SINTETIZADO

<400> 2

60 <400> 2

Pro Lys Lys Lys Arg Arg Val
1 5

<210> 3
 <211> 16
 <212> PRT
 5 <213> Secuencia artificial

 <220>
 <223> SINTETIZADO
 10 <400> 3

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

 15 <210> 4
 <211> 20
 <212> PRT
 <213> Secuencia artificial

 20 <220>
 <223> SINTETIZADO

 <400> 4

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

 Lys Arg Lys Val
 20
 25

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

 <220>
 <223> SINTETIZADO

 <400> 5
 35
 Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys
 1 5 10 15

 Arg Lys Val

 40 <210> 6
 <211> 24
 <212> PRT
 <213> Secuencia artificial

 <220>
 <223> SINTETIZADO
 45
 <400> 6

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

 Ala Pro Lys Lys Lys Arg Lys Val
 20

ES 2 713 243 T3

<210> 7
 <211> 27
 <212> PRT
 <213> Secuencia artificial
 5
 <220>
 <223> SINTETIZADO
 <400> 7
 10
 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
 <213> Secuencia artificial
 15
 <220>
 <223> SINTETIZADO
 20
 <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
 25
 <210> 9
 <211> 1374
 <212> PRT
 <213> Secuencia artificial
 30
 <220>
 <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
 35

ES 2 713 243 T3

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

ES 2 713 243 T3

Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	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 2 713 243 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

ES 2 713 243 T3

	820						825						830					
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							

ES 2 713 243 T3

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

ES 2 713 243 T3

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
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
<213> Secuencia artificial

<220>
<223> SINTETIZADO

<400> 10

atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	60
atcaccgacg actacaaggt gccagcaag aaattcaagg tgctgggcaa caccgaccgg	120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg gctctggcga aacagccgag	180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc	240
tatctgcaag agatcttcag caacgagatg gccaagggtg acgacagctt cttccacaga	300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcacc catcttcggc	360
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag	420
aagctggccg acagcaccga caaggccgac ctgagactga tctacctggc cctggcccac	480
atgatcaagt tccggggcca ctctctgac gagggcgacc tgaacccga caacagcgac	540
gtggacaagc tgttcatcca gctggtgcag atctacaatc agctgttcga ggaaaacccc	600
atcaacgcca gcagagtga cgccaaggcc atcctgagcg ccagactgag caagagcaga	660
cggctggaaa atctgatcg ccagctgccc ggcgagaagc ggaatggcct gttcggcaac	720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	780
gatgccaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc	900
ctgctgagcg acatcctgag agtgaacagc gagatcacca aggccccct gtccgcctct	960
atgatcaaga gatacgacga gcaccaccag gacctgacct tgctgaaagc tctcgtgcgg	1020

cagcagctgc	ctgagaagta	caaagagatt	ttcttcgacc	agagcaagaa	cggctacgcc	1080
ggctacatcg	atggcggagc	cagccaggaa	gagttctaca	agttcatcaa	gcccacctcg	1140
gaaaagatgg	acggcaccga	ggaactgctc	gtgaagctga	acagagagga	cctgctgcgg	1200
aagcagcggg	ccttcgacaa	cggcagcatc	ccccaccaga	tccacctggg	agagctgcac	1260
gccattctgc	ggcggcagga	agatttttac	ccattcctga	aggacaaccg	ggaaaagatc	1320
gagaagatcc	tgaccttcag	aatcccctac	tacgtgggcc	ctctggccag	gggaaacagc	1380
agattcgcct	ggatgaccag	aaagagcgag	gaaaccatca	ccccctggaa	cttcgaggaa	1440
gtggtggaca	agggcgccag	cgcccagagc	ttcatcgagc	ggatgaccaa	cttcgataag	1500
aacctgccc	acgagaaggt	gctgccc	cacagcctgc	tgtacgagta	cttcaccgtg	1560
tacaacgagc	tgaccaaagt	gaaatacgtg	accgagggaa	tgcggaagcc	cgcctttctg	1620
agcggcgagc	agaaaaaggc	catcgtggac	ctgctgttca	agaccaaccg	gaaagtgacc	1680
gtgaagcagc	tgaaagagga	ctacttcaag	aaaatcgagt	gcttcgacag	cgtggaaatc	1740
agcggcggtg	aagatcggtt	caacgcctcc	ctgggcgcct	atcacgatct	gctgaaaatt	1800
atcaaggaca	aggacttcct	ggacaatgag	gaaaacgagg	acattctgga	agatatcgtg	1860
ctgaccctga	cactgtttga	ggaccggggc	atgatcgagg	aacggctgaa	aacctatgcc	1920
cacctgttcg	acgacaaagt	gatgaagcag	ctgaagcggc	ggagatacac	cggctggggc	1980
aggctgagcc	ggaagctgat	caacggcatc	cgggacaagc	agtccggcaa	gacaatcctg	2040
gatttcctga	agtccgacgg	cttcgccaac	agaaacttca	tgcagctgat	ccacgacgac	2100
agcctgacct	ttaaagagga	catccagaaa	gcccaggtgt	ccggccaggg	acactctctg	2160
cacgagcaga	tcgccaatct	ggccggatcc	cccgccatta	agaagggc	cctgcagaca	2220
gtgaagattg	tggacgagct	cgtgaaagtg	atgggccc	agcccagagaa	catcgtgatc	2280
gaaatggcca	gagagaacca	gaccaccag	aagggacaga	agaacagccg	cagagagaatg	2340
aagcggatcg	aagagggcat	caaagagctg	ggcagccaga	tcctgaaaga	acaccccg	2400
gaaaacaccc	agctgcagaa	cgagaagctg	tacctgtact	acctgcagaa	tgggcgggat	2460
atgtacgtgg	accaggaact	ggacatcaac	cggctgtccg	actacgatgt	ggaccacatt	2520
gtgccccagt	ccttcatcaa	ggacgactcc	atcgataaca	aagtgctgac	tcggagcgac	2580
aagaaccggg	gcaagagcga	caacgtgccc	tccgaagagg	tcgtgaagaa	gatgaagaac	2640
tactggcgcc	agctgctgaa	tgccaagctg	attaccagaa	ggaagtctga	caatctgacc	2700
aaggccgaga	gaggcggcct	gagcgaactg	gataaggccg	gcttcattaa	gcggcagctg	2760
gtggaaaccc	ggcagatcac	aaagcacgtg	gcacagatcc	tggactccc	gatgaacact	2820
aagtacgacg	agaacgacaa	actgatccgg	gaagtgaag	tgatcaccct	gaagtccaag	2880

ES 2 713 243 T3

ctggtgtccg acttcagaaa ggatttccag ttttaciaaag tgcgcgagat caacaactac 2940
caccacgccc acgacgccta cctgaacgcc gtcgtgggaa ccgccctgat caaaaagtac 3000
cctaagctgg aaagcgagtt cgtgtacggc gattacaagg tgtacgacgt gcggaagatg 3060
atcgccaaga gcgagcagga aatcggcaag gctaccgcca agtacttctt ctacagcaac 3120
atcatgaact ttttcaagac cgagatcaca ctggccaacg gcgagatcag aaagcggcct 3180
ctgatcgaga caaacggcga aaccggggag atcgtgtggg ataagggccg ggattttgcc 3240
acagtgcgga aagtgtgtc catgccccaa gtgaatatcg tgaaaaagac cgaggtgcag 3300
accggcggct tcagcaaaga gtctatcctg cccaagagga actccgacaa gctgatcgcc 3360
agaaagaagg attgggaccc taagaagtac ggcggtttg acagccccac cgtggcctac 3420
tctgtgctgg tgggtggccaa agtggaaaag ggcaagtcca agaaactgaa gagtgtgaaa 3480
gagctgctgg ggatcaccat catggaaaaga agcagcttcg agaagaatcc catcgacttt 3540
ctggaagcca agggctacaa agaagtgaaa aaggacctga tcatcaagct gcctaagtac 3600
tcctgttctg agctggaaaa cggccggaag cggatgctgg cttctgccgg cgaactgcag 3660
aagggaaaac agctggccct gccctccaaa tatgtgaact tcctgtacct ggccagccac 3720
tatgagaagc tgaagggtc ccccgaggat aatgagcaga aacagctgtt tgtggaacag 3780
cacaagcact acctggacga gatcatcgag cagattagcg agttctccaa gcgcgtgatc 3840
ctggccgatg ccaacctgga caaggtgctg agcgcctaca acaagcaccg ggataagccc 3900
atcagagagc aggccgagaa tatcatccac ctgtttaccc tgaccaacct gggagcccct 3960
gccgccttca agtactttga caccaccatc gaccggaaga ggtacaccag caccaaagag 4020
gtgctggacg ccacctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac 4080
ctgtctcagc tgggaggcga cccaagaaa aagcgcaaag tg 4122

<210> 11
<211> 4764
<212> ADN
<213> *Homo sapiens*

<400> 11

gcggggcgggc ggtgcgatgt ccggagagga tggcccggcg gctggcccg gggcggcggc 60
ggcggctgcc cgggagcggc gacgggagca gctgcggcag tggggggcgc gggcggcgc 120
cgagcctggc cccggagagc gccgcgccc caccgtccgc ttcgagcgcg ccgccgagtt 180
cctggcggcc tgtgcgggcg gcgacctgga cgaggcgcgt ctgatgctgc gcgccgccga 240
ccctggcccc ggcgcgcgagc tcgacccgc cgcgcgcgcg cccgcccgcg ccgtgctgga 300
ctccaccaac gccgacggta tcagcgccct gcaccaggtc agcgcccccc gcccggcgtc 360
tcccggggcc aggtccaccc tctgctgcgc cacctggggc atcctccttc cccgttgcca 420

ES 2 713 243 T3

gtctcgatcc gccccgtcgt tcctggccct gggctttgcc accctatgct gacaccccg	480
cccagtcgcc cttaccattc cccttcgacc accccaacttc cgaattggag ccgcttcaac	540
tggccctggg cttagccact ctgtgctgac cactctgccc caggcctcct taccattccc	600
cttcgaccta ctctcttcg cattggagtc gctttaactg gccctggctt tggcagcctg	660
tgctgaccca tgcagtcctc cttaccatcc ctccctcgac ttcccctctt ccgatgttga	720
gccccctcag ccggtcctgg actttgtctc cttccctgcc ctgccctctc ctgaacctga	780
gccagctccc atagctcagt ctggtctatc tgcctggccc tggccattgt cactttgcgc	840
tgccctcctc tcgccccga gtgcccttgc tgtgccgcgc gaactctgcc ctctaacgct	900
gccgtctctc tcctgagtc ggaccacttt gagctctact ggcttctgcg ccgcctctgg	960
cccactgttt ccccttccca ggcaggtcct gctttctctg acctgcattc tctccctgg	1020
gcctgtgccg ctttctgtct gcagcttggt gcctgggtca cctctacggc tggcccagat	1080
ccttccctgc cgcctccttc aggttcctgc ttccctcaact cctcttccc cttgctctct	1140
gctgtgttgc tgcccaagga tgctctttcc ggagcacttc cttctcggcg ctgcaccacg	1200
tgatgtcctc tgagcggatc ctccccgtgt ctgggtcctc tcggggcatc tctcctccct	1260
cacccaaccc catgccgtct tcaactcgctg ggttcccttt tccttctcct tctggggcct	1320
gtgccatctc tcgtttctta ggatggcctt ctccgacgga tgtctccctt gcgtcccgcc	1380
tccccctctt gtaggcctgc atcatcacgc tttttctgga caaccccaaa gtaccccgctc	1440
tccctggctt tagccacctc tccatcctct tgctttcttt gcctggacac cccgttctcc	1500
tgtggattcg ggtaacctct cactcctttc atttgggcag ctcccctacc ccccttacct	1560
ctctagtctg tgctagtctt tccagcccc tgatcatggca tcttccaggg gtccgagagc	1620
tcagctagtc ttcttctctc aaccggggcc cctatgtcca cttcaggaca gcatgtttgc	1680
tgcctccagg gatcctgtgt ccccgagctg ggaccacctt atattcccag ggccgggttaa	1740
tgtggctctg gttctgggtg cttttatctg tcccctccac cccacagtgg ggccactagg	1800
gacaggattg gtgacagaaa agcccatcc ttaggcctcc tccttcttag tctcctgata	1860
ttgggtctaa cccccacctc ctgttaggca gattccttat ctggtgacac accccattt	1920
cctggagcca tctctctcct tgccagaacc tctaagggtt gcttacgatg gagccagaga	1980
ggatcctggg agggagagct tggcaggggg tgggagggaa gggggggatg cgtgacctgc	2040
ccggttctca gtggccaccc tgcgctaccc tctcccagaa cctgagctgc tctgacgcgg	2100
ccgtctggtg cgtttctact atcctggtgc tgcagcttcc ttacacttcc caagaggaga	2160
agcagtttgg aaaaacaaaa tcagaataag ttggtcctga gttctaactt tggctcttca	2220
cctttctagt cccaatttta tattgttctt ccgtgcgtca gttttacctg tgagataagg	2280
ccagtagcca gccccgtcct ggcagggctg tggtagaggag ggggggtgtcc gtgtggaaaa	2340

ES 2 713 243 T3

ctccctttgt	gagaatggtg	cgtcctaggt	gttcaccagg	tcgtggccgc	ctctactccc	2400
tttctctttc	tccatccttc	tttccttaaa	gagtccccag	tgctatctgg	gacatattcc	2460
tccgcccaga	gcagggtccc	gcttccctaa	ggccctgctc	tgggcttctg	ggtttgagtc	2520
cttggaagc	ccaggagagg	cgctcaggct	tccctgtccc	ccttcctcgt	ccaccatctc	2580
atgcccctgg	ctctcctgcc	ccttccctac	aggggttcct	ggctctgctc	ttcagactga	2640
gccccgttcc	cctgcatccc	cgttcccctg	catccccctt	cccctgcatc	ccccagaggc	2700
cccaggccac	ctacttggcc	tggaccccac	gagaggccac	cccagccctg	tctaccaggc	2760
tgccttttgg	gtggattctc	ctccaactgt	ggggtgactg	cttggaacac	tcactcttcg	2820
gggtatccca	ggaggcctgg	agcattgggg	tgggctgggg	ttcagagagg	agggattccc	2880
ttctcaggtt	acgtggccaa	gaagcagggg	agctgggttt	gggtcaggtc	tgggtgtggg	2940
gtgaccagct	tatgctgttt	gccaggaca	gcctagtttt	agcactgaaa	ccctcagtcc	3000
taggaaaaca	gggatggttg	gtcactgtct	ctgggtgact	cttgattccc	ggccagtttc	3060
tccacctggg	gctgtgtttc	tcgtcctgca	tccttctcca	ggcagggtccc	caagcatcgc	3120
ccccctgctg	tggctgttcc	caagtcttta	gggtacccca	cgtgggttta	tcaaccactt	3180
ggtgaggctg	gtaccctgcc	cccatcctg	cacccaatt	gccttagtgg	ctaggggggt	3240
gggggctaga	gtaggagggg	ctggagccag	gattcttagg	gctgaacaga	gaagagctgg	3300
gggcctgggc	tcctgggttt	gagagaggag	gggctggggc	ctggactcct	gggtccgagg	3360
gaggaggggc	tggggcctgg	actcctgggt	ctgagggtgg	agggactggg	ggcctggact	3420
cctgggtccg	agggaggagg	ggctggggcc	tggactcgtg	ggtctgaggg	aggaggggct	3480
gggggcctgg	acttctgggt	cttagggagg	cggggctggg	cctggacccc	tgggtctgaa	3540
tggggagagg	ctgggggcct	ggactccttc	atctgagggc	ggaagggctg	gggcctggcc	3600
tcctgggttg	aatggggagg	ggttgggcct	ggactctgga	gtccctgggtg	cccaggcctc	3660
aggcatcttt	cacagggatg	cctgtactgg	gcaggtcctt	gaaagggaaa	ggccatttgc	3720
tctccttgcc	cccctcccct	atcgccatga	caactgggtg	gaaataaacg	agccgagttc	3780
atcccgttcc	cagggcacgt	gcggcccctt	cacagcccga	gtttccatga	cctcatgctc	3840
ttggccctcg	tagctccctc	ccgcctcctc	cagatgggca	gctttggaga	ggtgagggac	3900
ttggggggta	atttatcccg	tggatctagg	agtttagctt	cactccttcc	tcagctccag	3960
ttcagggtccc	ggagcccacc	cagtgtccac	aaggcctggg	gcaagtccct	cctccgaccc	4020
cctggacttc	ggcttttgtc	cccccaagtt	ttggacccct	aagggaagaa	tgagaaacgg	4080
tggcccgtgt	cagcccctgg	ctgcagggcc	ccgtgcagag	ggggcctcag	tgaactggag	4140
tgtgacagcc	tggggcccag	gcacacaggt	gtgcagctgt	ctcacccttc	tgggagttcc	4200

ES 2 713 243 T3

	gcccaggccc ctgagtctgt cccagcacag ggtggccttc ctccaccctg catagccctg	4260
	ggcccacggc ttcgttcctg cagagtatct gctgggggtgg tttccgagct tgacccttgg	4320
	aaggacctgg ctgggtttta ggcaggagg gctggggggcc aggactcctg gctctgaagg	4380
	aggaggggct ggaacctctt ccctagtctg agcactggaa gcgccacctg tgggtggtga	4440
	cggggggttt gccgtgtcta acaggtagcca tgtgggggttc ccgcaccag atgagaagcc	4500
	ccctcccttc cccgttcact tcctgtttgc agatagccag gagtcctttc gtgggtttcca	4560
	ctgagcactg aaggcctggc cggcctgacc actgggcaac caggcgtatc ttaaacagcc	4620
	agtggccaga ggctgttggg tcattttccc cactgtccta gcaccgtgtc cctggatctg	4680
	ttttcgtggc tccctctgga gtcccgactt gctgggacac cgtggctggg gtaggtgcgg	4740
	ctgacggctg tttcccaccc ccag	4764
5	<210> 12 <211> 42 <212> ARN <213> Secuencia artificial	
10	<220> <223> SINTETIZADO	
	<400> 12 accccacagu ggggccacua guuuuagagc uaugcuguuu ug	42
15	<210> 13 <211> 86 <212> ARN <213> Secuencia artificial	
20	<220> <223> SINTETIZADO	
	<400> 13	
25	ggaaccuuc aaaacagcau agcaaguuaa aaauaggcua guccguuau aacuugaaaa	60
	aguggcaccg agucggugcu uuuuuu	86
30	<210> 14 <211> 62 <212> ARN <213> Secuencia artificial	
	<220> <223> SINTETIZADO	
35	<400> 14 accccacagu ggggccacua guuuuagagc uagaaaauagc aaguuaaaau aaggcuaguc	60
	cg	62
40	<210> 15 <211> 25 <212> ADN <213> Secuencia artificial	

ES 2 713 243 T3

<220>
 <223> SINTETIZADO

5 <400> 15
 ctgaccttt ctctctctcc cacag 25

10 <210> 16
 <211> 1009
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> SINTETIZADO

<400> 16

gccacccatgg actacaaaga cgatgacgac aaggctcgact ctagagctgc agagagcgac	60
gagagcgggcc tgcccgccat ggagatcgag tgccgcatca ccggcaccct gaacggcgtg	120
gagttcgagc tgggtgggcgg cggagagggc acccccgagc agggccgcat gaccaacaag	180
atgaagagca ccaaaggcgc cctgaccttc agcccctacc tgctgagcca cgtgatgggc	240
tacggcttct accacttcgg cacctacccc agcggctacg agaaccctt cctgcacgcc	300
atcaacaacg gcggtctacac caacaccgc atcgagaagt acgaggacgg cggcgtgctg	360
cacgtgagct tcagctaccg ctacgaggcc ggccgcgtga tcggcgactt caaggtgatg	420
ggcaccggct tccccgagga cagcgtgatc ttcaccgaca agatcgtccg cagcaacgcc	480
accgtggagc acctgcaccc catgggcat aacgatctgg atggcagctt caccgcacc	540
ttcagcctgc gcgacggcgg ctactacagc tccgtggtgg acagccacat gcacttcaag	600
agcgccatcc accccagcat cctgcagaac gggggcccca tggtcgctt ccgccgcgtg	660
gaggaggatc acagcaacac cgagctgggc atcgtagagt accagcacgc cttcaagacc	720
ccggatgcag atgccggtga agaataaga tctctgtgcc ttctagttgc cagccatctg	780
ttgtttgccc ctcccccggtg ccttccttga ccctggaagg tgccactccc actgtccttt	840
cctaataaaa tgaggaaatt gcatcgcat gtctgagtag gtgtcattct attctggggg	900
gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg	960
atgcggtggg ctctatggac tcgaggttta aacgtcgacg cggccgcgt	1009

20 <210> 17
 <211> 355
 <212> PRT
 <213> Secuencia artificial

25 <220>
 <223> SINTETIZADO

<400> 17

ES 2 713 243 T3

Met	Ser	Gly	Glu	Asp	Gly	Pro	Ala	Ala	Gly	Pro	Gly	Ala	Ala	Ala	Ala	
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	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 2 713 243 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

5 <210> 18
<211> 21
<212> ADN
<213> Secuencia artificial

10 <220>
<223> SINTETIZADO

<400> 18
ccactctgtg ctgaccactc t 21

15 <210> 19
<211> 17
<212> ADN
<213> Secuencia artificial

20 <220>
<223> SINTETIZADO

<400> 19
gcggcactcg atctcca 17

25 <210> 20
<211> 711
<212> AND
<213> *Mus musculus*

30 <400> 20

ES 2 713 243 T3

gagcggctgc ggggcgggtg caagcacgtt tccgacttga gttgcctcaa gaggggcgtg 60
ctgagccaga cctccatcgc gcactccggg gagtggaggg aaggagcgag ggctcagttg 120
ggctgttttg gaggcaggaa gcacttgctc tcccaaagtc gctctgagtt gttatcagta 180
agggagctgc agtggagtag gcggggagaa ggccgcaccc ttctccggag gggggagggg 240
agtgttgcaa tacctttctg ggagttctct gctgcctcct ggcttctgag gaccgccctg 300
ggcctgggag aatcccttcc ccctcttccc tcgtgatctg caactccagt ctttctagaa 360
gatgggcggg agtcttctgg gcaggcttaa aggctaacct ggtgtgtggg cgttgtcctg 420
caggggaatt gaacaggtgt aaaattggag ggacaagact tcccacagat tttcggtttt 480
gtcgggaagt tttttaatag gggcaaataa ggaaaatggg aggataggta gtcactctggg 540
gttttatgca gcaaaactac aggttattat tgcttgtgat ccgcctcgga gtattttcca 600
tcgaggtaga ttaaagacat gtcacccga gttttatact ctctgcttg agatccttac 660
tacagtatga aattacagtg tcgcgagtta gactatgtaa gcagaatttt a 711

<210> 21
<211> 42
<212> ARN
<213> Secuencia artificial

<220>
<223> SINTETIZADO

<400> 21
cuccagucuu ucuagaagau guuuuagagc uaugcuguuu ug 42

<210> 22
<211> 42
<212> ARN
<213> Secuencia artificial

<220>
<223> SINTETIZADO

<400> 22
ugaacaggug uaaaauugga guuuuagagc uaugcuguuu ug 42

<210> 23
<211> 42
<212> ARN
<213> Secuencia artificial

<220>
<223> SINTETIZADO

<400> 23
ugucgggaag uuuuuuaaua guuuuagagc uaugcuguuu ug 42

<210> 24
<211> 642
<212> ADN
<213> *Rattus rattus*

ES 2 713 243 T3

<400> 24

gggattcctc cttgagttgt ggcactgagg aacgtgctga acaagaccta cattgcactc	60
cagggagtggt atgaaggagt tggggctcag tcgggttgta ttggagacaa gaagcacttg	120
ctctccaaaa gtcggtttga gttatcatta agggagctgc agtggagtag gcggagaaaa	180
ggccgcaccc ttctcaggac gggggagggg agtggtgcaa tacctttctg ggagttctct	240
gctgcctcct gtcttctgag gaccgccctg ggcctggaag attcccttcc cccttcttcc	300
ctcgtgatct gcaactggag tctttctgga agataggcgg gagtcttctg ggcaggctta	360
aaggctaacc tgggtgcgtgg ggcgttgctc tgcagaggaa ttgaacaggt gtaaaattgg	420
aggggcaaga cttccacag attttcgatt gtgttgtaa gtattgtaat aggggcaa	480
aagggaata gactaggcac tcacctggg ttttatgcag caaaactaca gggtattatt	540
gcttgatgatc cgccctggag aatttttcac cgaggtagat tgaagacatg cccacc	600
caaa ttttaatat cttccacttg cgatccttgc tacagtatga aa	642

5 <210> 25
<211> 42
<212> ARN
<213> Secuencia artificial

10 <220>
<223> SINTETIZADO

<400> 25
agggggaagg gaauccuucca guuuuagagc uaugcuguuu ug 42

15 <210> 26
<211> 42
<212> ARN
<213> Secuencia artificial

20 <220>
<223> SINTETIZADO

<400> 26
ucugcaacug gagucuuuc guuuuagagc uaugcuguuu ug 42

25 <210> 27
<211> 42
<212> ARN
<213> Secuencia artificial

30 <220>
<223> SINTETIZADO

35 <400> 27
aggcgggagu cuucugggca guuuuagagc uaugcuguuu ug 42

REIVINDICACIONES

1. Un procedimiento para modificar una secuencia cromosómica en una célula eucariota, comprendiendo el procedimiento:

a) Introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas, y, opcionalmente, un polinucleótido donador, en el que cada sistema de nickasa guiado por ARN comprende

(i) Una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble; y

(ii) Un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interactúa con la endonucleasa guiada por ARN,

en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM), y los sitios diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y

b) Cultivar la célula eucariótica de manera que las dos endonucleasas guiadas por ARN escinden cadenas opuestas de la secuencia cromosómica en estrecha cercanía, lo suficiente como para introducir una rotura de cadena doble en la secuencia cromosómica, y reparar la rotura de cadena doble mediante un proceso de reparación de ADN que lleva a la modificación de la secuencia cromosómica, en la que

el procedimiento 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.

2. Un procedimiento ex vivo o in vitro para modificar una secuencia cromosómica en una célula eucariota, comprendiendo el procedimiento:

a) introducir en la célula eucariota dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas, y, opcionalmente, un polinucleótido donador, en el que cada sistema de nickasa guiado por ARN comprende

(i) una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble; y

(ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en la secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN,

en el que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM) y los sitios diana de las dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica; y

b) cultivar la célula eucariota, de manera que las dos endonucleasas guiadas por ARN escinden cadenas opuestas de la secuencia cromosómica en estrecha cercanía, lo suficiente como para introducir una rotura de cadena doble en la secuencia cromosómica, y reparar la rotura de cadena doble mediante un proceso de reparación de ADN que lleva a la modificación de la secuencia cromosómica, en el que

el procedimiento no comprende un procedimiento para modificar la identidad genética de la línea germinal de un ser humano.

3. El procedimiento de la reivindicación 1 o de la reivindicación 2, en el que cada endonucleasa guiada por ARN deriva de repeticiones palindrómicas cortas agrupadas y espaciadas de forma regular (CRISPR)/sistema proteico de (Cas) asociado a CRISPR (CRISPR/Cas) de tipo II.

4. El procedimiento de la reivindicación 3, en el que la proteína del sistema CRISPR/Cas de tipo II es una proteína Cas9.

5. El procedimiento de la reivindicación 4, en el que la proteína Cas9 comprende una mutación en el dominio RuvC o HNH.

6. El procedimiento de cualquier reivindicación previa, en el que cada endonucleasa guiada por ARN comprende adicionalmente al menos un dominio adicional elegido de una señal de localización nuclear, un dominio de penetración celular, o un dominio marcador.

7. El procedimiento de cualquier reivindicación previa, en el que el sitio diana es un locus Rosa26, un locus HPRT o un locus AAVS1.

8. El procedimiento de cualquier reivindicación previa, en el que el polinucleótido donador comprende una secuencia donadora que tiene al menos un cambio de nucleótido en relación con la secuencia cromosómica cerca de los sitios diana en la secuencia cromosómica.

9. El procedimiento de la reivindicación 8, en el que la secuencia donadora está flanqueada por secuencias que tienen identidad de secuencia sustancial con secuencias localizadas aguas arriba y aguas abajo del sitio diana en la secuencia cromosómica.
- 5 10. El procedimiento de la reivindicación 8 o de la reivindicación 9, en el que la secuencia donadora está flanqueada por proyecciones cortas que son compatibles con proyecciones generadas por la endonucleasa guiada por ARN.
11. El procedimiento de cualquier reivindicación previa, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ARNm y el ácido nucleico que codifica cada ARN guía es ADN.
12. El procedimiento de cualquiera de las reivindicaciones 1-10, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ADN y el ácido nucleico que codifica cada ARN guía es ADN.
- 10 13. El procedimiento de la reivindicación 12, en el que el ADN es parte de un vector que comprende adicionalmente una secuencia de control del promotor que está operativamente unida al ácido nucleico que codifica la endonucleasa guiada por ARN y una secuencia de control del promotor que está operativamente unida al ácido nucleico que codifica el ARN guía.
- 15 14. El procedimiento de cualquiera de las reivindicaciones 1-13, en el que la célula eucariota es una célula humana, una célula de mamífero no humano, o un embrión de mamífero no humano.
- 15 15. El procedimiento de cualquiera de las reivindicaciones 1-13, en el que la célula eucariota es una célula de invertebrado, una célula de insecto, una célula de planta, una célula de levadura o un organismo eucariota unicelular.
16. El procedimiento de la reivindicación 15, en el que la célula eucariota es una célula de planta.
- 20 17. El procedimiento de la reivindicación 14, en el que la célula humana o la célula de mamífero no humano es un linfocito T.
18. El procedimiento de cualquiera de las reivindicaciones 1-17, en el que la célula eucariota está *in vitro*.
19. El procedimiento de cualquier reivindicación previa, en el que el polinucleótido donador no se introduce en la célula eucariota, y la reparación de la rotura de cadena doble mediante un proceso de reparación de unión de extremos no homólogos da como resultado la inactivación de la secuencia cromosómica.
- 25 20. El procedimiento de cualquiera de las reivindicaciones 1-18, en el que el polinucleótido donador se introduce en la célula eucariota, y la reparación de la rotura de cadena doble da como resultado un cambio de al menos un nucleótido en la secuencia cromosómica.
21. El procedimiento de la reivindicación 20, en el que el cambio comprende una integración de una secuencia exógena.
- 30 22. El procedimiento de una cualquiera de las reivindicaciones 1-10 o 14-21, en el que el al menos uno ARN guía se sintetiza químicamente.
23. Una composición que comprende al menos dos sistemas de nickasa guiados por ARN, comprendiendo cada sistema de nickasa guiado por ARN (i) una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble, y (ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en una secuencia cromosómica y una segunda región que interactúa con la endonucleasa guiada por ARN, en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM), y los sitios diana de dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica y en estrecha cercanía, lo suficiente como para introducir una rotura de cadena doble en la secuencia cromosómica.
- 40 24. La composición de la reivindicación 23, en la que cada endonucleasa guiada por ARN deriva de repeticiones palindrómicas cortas agrupadas y espaciadas de forma regular (CRISPR)/sistema proteico de (Cas) asociado a CRISPR (CRISPR/Cas) de tipo II.
25. La composición de la reivindicación 24, en la que la proteína del sistema CRISPR/Cas de tipo II es una proteína Cas9.
- 45 26. La composición de la reivindicación 25, en la que la proteína Cas9 comprende una mutación en el dominio RuvC o HNH.
27. La composición de cualquiera de las reivindicaciones 23-26, en la que cada endonucleasa guiada por ARN comprende adicionalmente al menos un dominio adicional elegido de una señal de localización nuclear, un dominio de penetración celular, o un dominio marcador.
- 50

28. La composición de cualquiera de las reivindicaciones 23-27, en la que el sitio diana es un locus Rosa26, un locus HPRT, o un locus AAVS1.
29. La composición de cualquiera de las reivindicaciones 23-28, en la que el al menos uno ARN guía se sintetiza químicamente.
- 5 30. Un ácido nucleico que codifica la composición de cualquiera de las reivindicaciones 23-28.
31. El ácido nucleico de la reivindicación 30, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ARNm y el ácido nucleico que codifica cada ARN guía es ADN.
32. El ácido nucleico de la reivindicación 30, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ADN y el ácido nucleico que codifica cada ARN guía es ADN.
- 10 33. Un vector que comprende el ácido nucleico de la reivindicación 32, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN está unido de manera operativa a una secuencia de control del promotor para la expresión en una célula eucariota, y el ácido nucleico que codifica cada ARN guía está unido de forma operativa a una secuencia de control del promotor para la expresión en una célula eucariota.
- 15 34. Un kit que comprende al menos dos sistemas de nickasa guiados por ARN o ácido nucleico que codifica dichos sistemas, comprendiendo cada sistema de nickasa guiado por ARN (i) una endonucleasa guiada por ARN que se modifica para escindir una cadena de una secuencia de cadena doble y (ii) un ARN guía que comprende una primera región que tiene complementariedad con un sitio diana en una secuencia cromosómica y una segunda región que interacciona con la endonucleasa guiada por ARN, en la que cada sitio diana está seguido inmediatamente por un motivo adyacente de protoespaciador (PAM) y los sitios diana de dos endonucleasas guiadas por ARN están en cadenas opuestas de la secuencia cromosómica y en estrecha cercanía, lo suficiente como para introducir una rotura de cadena doble en la secuencia cromosómica.
- 20 35. El kit de la reivindicación 34, que comprende adicionalmente al menos un polinucleótido donador.
36. El kit de la reivindicación 34 o la reivindicación 35, en la que cada endonucleasa guiada por ARN deriva de repeticiones palindrómicas cortas agrupadas y espaciadas de forma regular (CRISPR)/sistema proteico de (Cas) asociado a CRISPR (CRISPR/Cas) de tipo II.
- 25 37. El kit de la reivindicación 36, en el que la proteína del sistema CRISPR/Cas de tipo II es una proteína Cas9.
38. El kit de la reivindicación 37, en el que la proteína Cas9 comprende una mutación en el dominio RvC o HNH.
39. El kit de cualquiera de las reivindicaciones 34-38, en el que cada endonucleasa guiada por ARN comprende adicionalmente al menos un dominio adicional elegido de una señal de localización nuclear, un dominio de penetración celular, o un dominio marcador.
- 30 40. El kit de cualquiera de las reivindicaciones 34-39, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ARNm y el ácido nucleico que codifica cada ARN guía es ADN.
41. El kit de cualquiera de las reivindicaciones 34-39, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN es ADN y el ácido nucleico que codifica cada ARN guía es ADN.
- 35 42. El kit de la reivindicación 41, en el que el ácido nucleico que codifica cada endonucleasa guiada por ARN está unido de manera operativa a una secuencia de control de promotor para la expresión en una célula eucariota, y el ácido nucleico que codifica cada ARN guía está unido de manera operativa a una secuencia de control del promotor para la expresión en una célula eucariota.
43. El kit de la reivindicación 35, en el que el polinucleótido donador comprende una secuencia donadora.
- 40 44. El kit de la reivindicación 43, en el que la secuencia donadora está flanqueada por secuencias que tienen identidad de secuencia sustancial con secuencias en cualquiera de los lados de los sitios diana en la secuencia cromosómica.
45. El kit de la reivindicación 43, en el que la secuencia donadora está flanqueada por proyecciones cortas que son compatibles con proyecciones generadas por la endonucleasa guiada por ARN.
- 45 46. El kit de cualquiera de las reivindicaciones 34-39 o 43-45, en el que el al menos uno ARN guía se sintetiza químicamente.

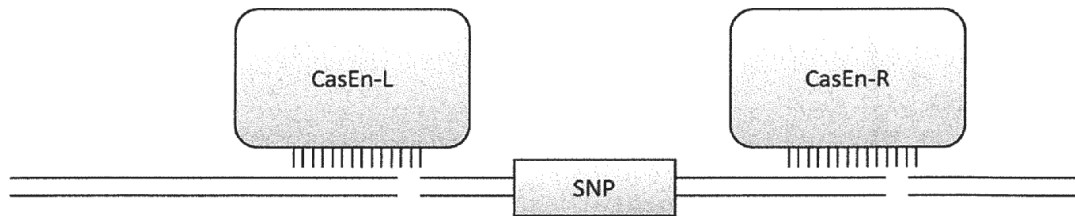


FIG. 1

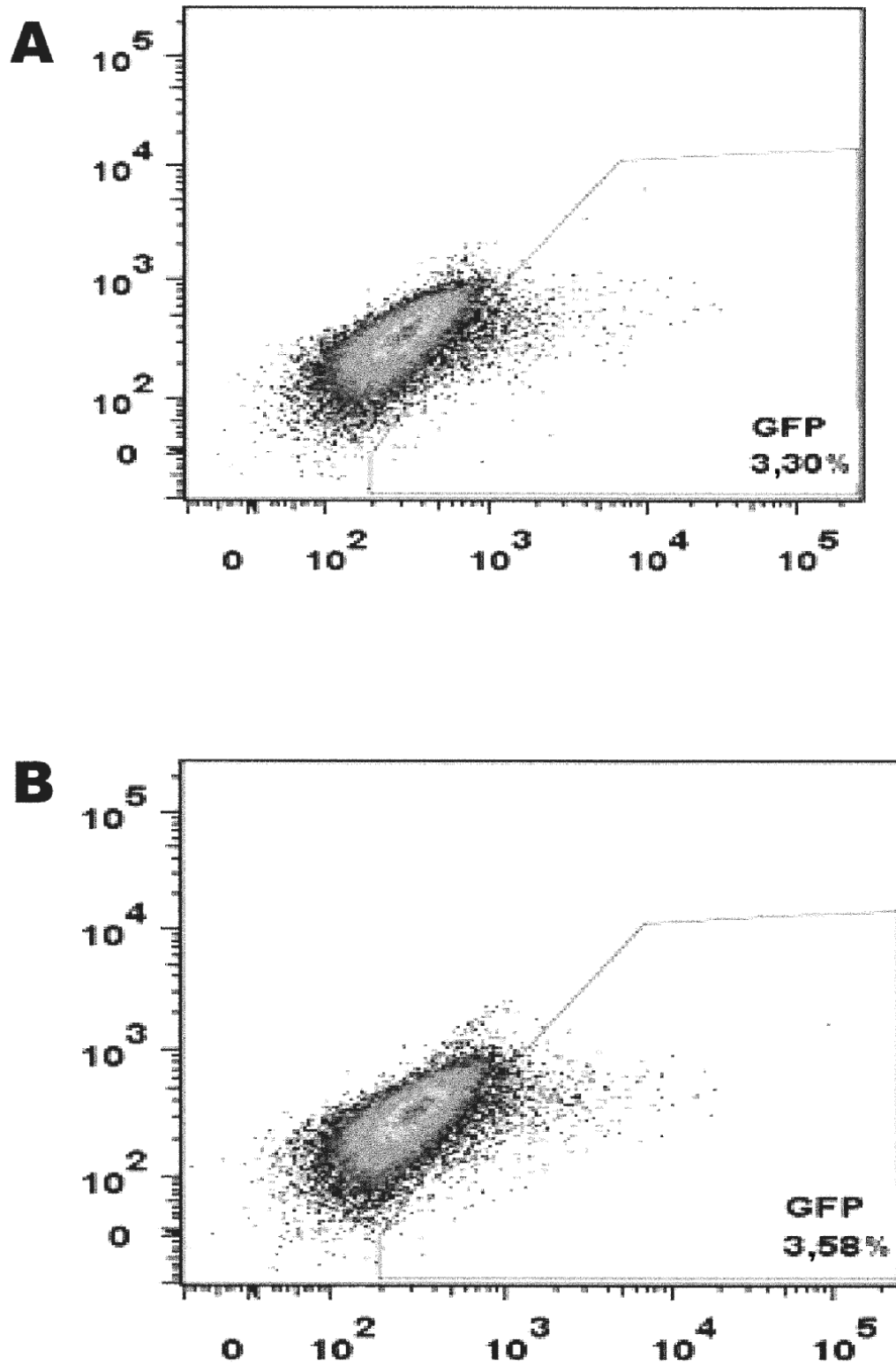


FIG. 2

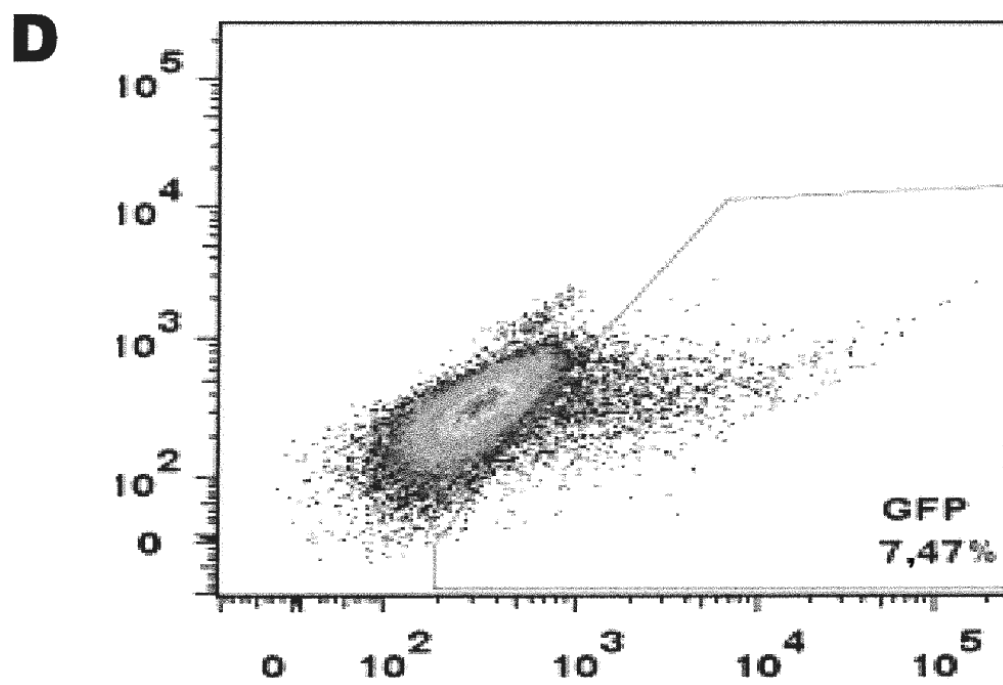
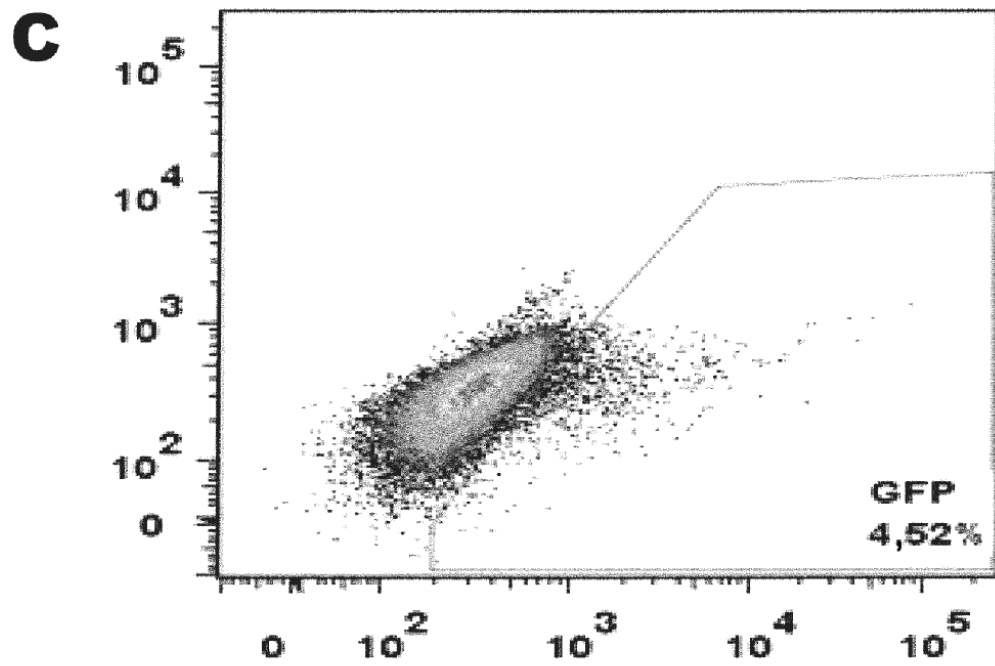
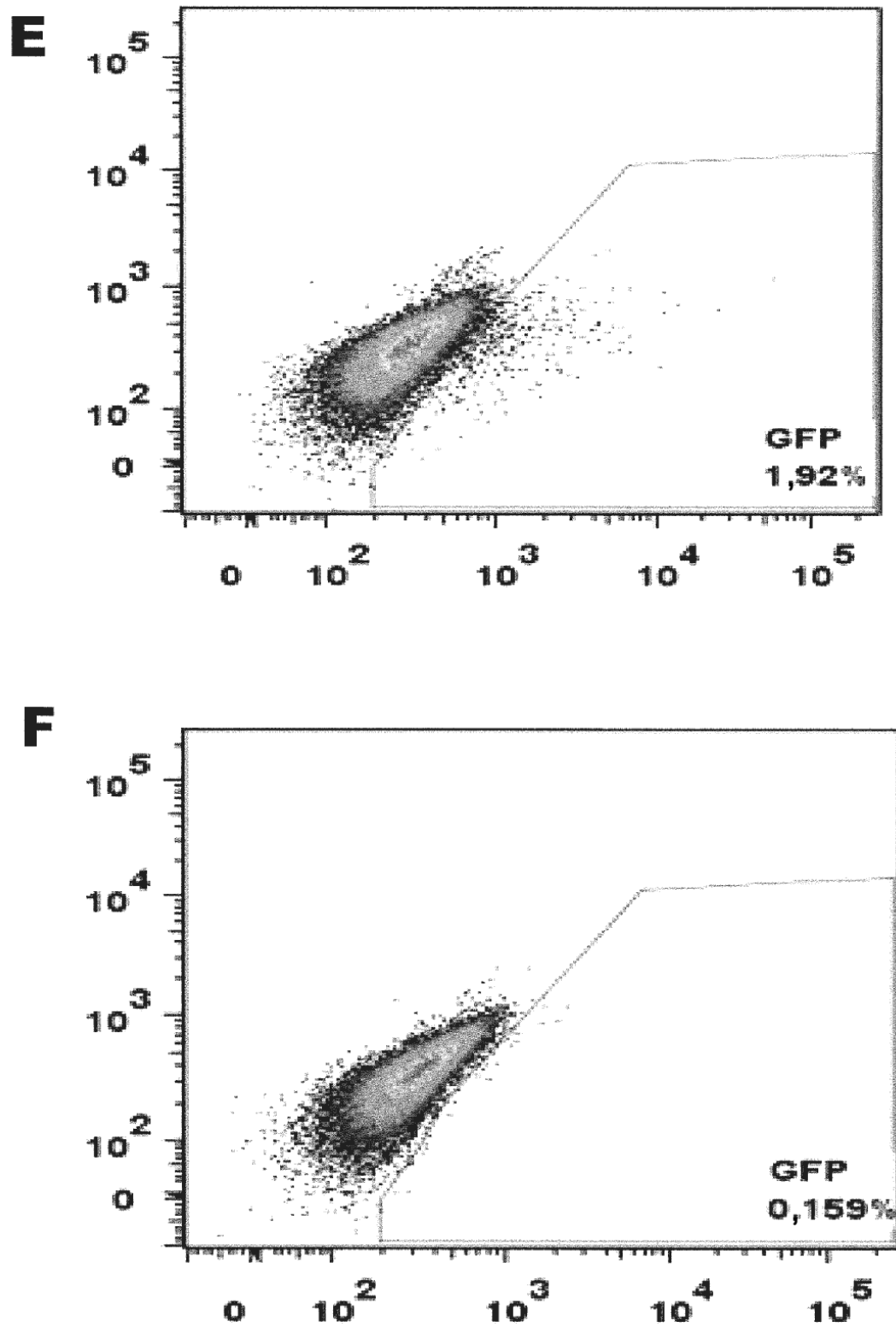


FIG. 2



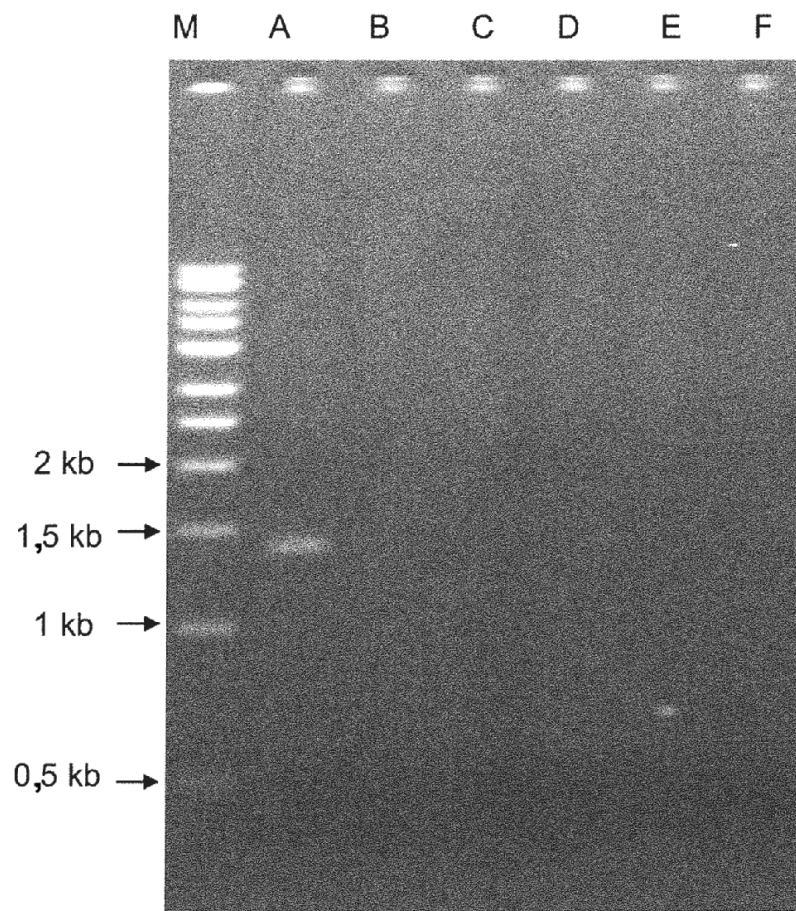


FIG. 3