

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **2 987 399**

(51) Int. Cl.:

A61K 48/00 (2006.01)
C12N 15/864 (2006.01)
C12N 15/90 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Fecha de presentación y número de la solicitud europea: **05.06.2014 E 19168481 (0)**

(97) Fecha y número de publicación de la concesión europea: **14.02.2024 EP 3539573**

(54) Título: **Edición génica guiada por ARN y regulación génica**

(30) Prioridad:

05.06.2013 US 201361831481 P
25.06.2013 US 201361839127 P
15.11.2013 US 201361904911 P
19.03.2014 US 201461967466 P
18.04.2014 US 201461981575 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.11.2024

(73) Titular/es:

DUKE UNIVERSITY (100.0%)
2812 Erwin Road, Suite 406
Durham, NC 27705, US

(72) Inventor/es:

GERSBACH, CHARLES;
HILTON, ISAAC B.;
PEREZ-PINERA, PABLO;
KABADI, AMI M.;
THAKORE, PRATIKSHA I.;
OUSTEROUT, DAVID y
BLACK, JOSHUA B.

(74) Agente/Representante:

MILTENYI, Peter

Observaciones:

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

ES 2 987 399 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

Edición génica guiada por ARN y regulación génica

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica prioridad sobre la Solicitud Provisional de EE.UU. N.º 61/831.481, presentada el 5 de junio de 2013, la solicitud provisional de EE.UU. N.º 61/839.127, presentada el 25 de junio de 2013, la solicitud provisional de

- 5 EE.UU. N.º 61/904.911, presentada el 15 de noviembre de 2013, la solicitud provisional de EE.UU. N.º 61/967.466, presentada el 19 de marzo de 2014 y la Solicitud Provisional de EE.UU. N.º 61/981.575, presentada el 18 de abril de 2014.

Declaración de interés gubernamental

- 10 Esta invención se realizó con el apoyo del gobierno bajo los números de subvención federal DP2-OD008586 y R01DA036865 otorgados por N1H y CBET-1151035 otorgado por la National Science Foundation. El gobierno de los Estados Unidos tiene ciertos derechos sobre esta invención.

Campo técnico

- 15 La presente divulgación se refiere al campo de la alteración de la expresión génica, la ingeniería genómica y la alteración genómica de genes usando sistemas basados en Grupos de Repeticiones Palindrómicas Cortas en Intervalos Regulares (CRISPR)/asociados a CRISPR (Cas) 9 y sistemas de suministro vírico. La presente divulgación 20 también se refiere al campo de la ingeniería genómica y la alteración genómica de genes en músculo, tales como músculo esquelético y músculo cardíaco.

Antecedentes

- 25 Se han diseñado factores de transcripción sintéticos para controlar la expresión génica para muchas aplicaciones médicas y científicas diferentes en sistemas de mamíferos, incluyendo estimulación de la regeneración de tejidos, detección de fármacos, compensación de defectos genéticos, activación de supresores de tumores silenciados, control de la diferenciación de células madre, realización de análisis genéticos y creación de circuitos de genes sintéticos. Estos factores de transcripción pueden dirigirse a promotores o potenciadores de genes endógenos, o diseñarse 30 específicamente para reconocer secuencias ortogonales a los genomas de mamíferos para la regulación de transgenes. Las estrategias más comunes para diseñar novedosos factores de transcripción dirigidos a secuencias definidas por el usuario se han basado en los dominios programables de unión al ADN de las proteínas con dedos de cinc y los efectores similares a los activadores de la transcripción (TALE). Ambos enfoques implican la aplicación de 35 los principios de las interacciones proteína-ADN de estos dominios para diseñar nuevas proteínas con una especificidad única de unión al ADN. Aunque estos métodos han tenido mucho éxito en muchas aplicaciones, la ingeniería de proteínas necesaria para manipular las interacciones proteína-ADN puede ser laboriosa y requerir experiencia especializada.

- 40 Adicionalmente, estas nuevas proteínas no siempre son eficaces. Las razones de esto aún no se conocen pero pueden estar relacionadas con los efectos de las modificaciones epigenéticas y el estado de la cromatina en la unión de las proteínas al sitio diana genómico. Además, existen desafíos para garantizar que estas nuevas proteínas, así como otros componentes, se suministren a cada célula. Los métodos existentes para suministrar estas nuevas proteínas y sus múltiples componentes incluyen el suministro a células en plásmidos o vectores separados, lo que conduce a niveles de expresión muy variables en cada célula debido a diferencias en el número de copias. Adicionalmente, la 45 activación génica después de la transfección es transitoria debido a la dilución del ADN plasmídico y la expresión génica temporal puede no ser suficiente para inducir efectos terapéuticos. Adicionalmente, este enfoque no es adecuado para tipos celulares que no se transfectan fácilmente. Por lo tanto, otra limitación de estas nuevas proteínas es la potencia de la activación transcripcional.

- 50 Pueden usarse nucleasas específicas de sitio para introducir roturas bicatenarias específicas de sitio en loci genómicos marcados como diana. Esta escisión del ADN estimula la maquinaria natural de reparación del ADN, dando lugar a una de las dos posibles vías de reparación. En ausencia de una plantilla donadora, la rotura se reparará mediante unión de extremos no homólogos (Non-Homologous End Joining, NHEJ), una vía de reparación propensa a errores que conduce a pequeñas inserciones o eliminaciones de ADN. Este método puede usarse para interrumpir, eliminar o alterar intencionalmente el marco de lectura de secuencias génicas específicas. Sin embargo, si se proporciona una plantilla donadora junto con las nucleasas, entonces la maquinaria celular reparará la rotura mediante recombinación homóloga, que se mejora varios órdenes de magnitud en presencia de escisión del ADN. Este método puede usarse 55 para introducir cambios específicos en la secuencia de ADN en sitios diana. Se han usado nucleasas diseñadas para la edición génica en una diversidad de líneas celulares y células madre humanas y para la edición génica en el hígado de ratón. Sin embargo, el principal obstáculo para la implementación de estas tecnologías es su suministro a tejidos concretos *in vivo* de una manera que sea eficaz, eficiente y facilite la modificación exitosa del genoma.

Las enfermedades genéticas hereditarias tienen efectos devastadores en los niños de los Estados Unidos. Actualmente estas enfermedades no tienen cura y sólo pueden controlarse intentando aliviar los síntomas. Durante décadas, el campo de la terapia génica ha prometido una cura para estas enfermedades. Sin embargo, los obstáculos técnicos relacionados con el suministro seguro y eficiente de genes terapéuticos a las células y a los pacientes han limitado este enfoque. La distrofia muscular de Duchenne (DMD) es la enfermedad monogénica hereditaria más común y se produce en 1 de cada 3500 hombres. La DMD es el resultado de mutaciones hereditarias o espontáneas en el gen de la distrofina. La distrofina es un componente clave de un complejo proteico que es responsable de regular la integridad y la función de las células musculares. Los pacientes con DMD normalmente pierden la capacidad de mantenerse físicamente a sí mismos durante la niñez, debilitándose progresivamente durante la adolescencia y mueren a los veinte años. Las estrategias actuales de terapia génica experimental para la DMD requieren la administración repetida de vehículos de suministro de genes transitorios o dependen de la integración permanente de material genético extraño en el ADN genómico. Ambos métodos tienen graves problemas de seguridad. Adicionalmente, estas estrategias se han visto limitadas por la incapacidad de suministrar la gran y compleja secuencia del gen de la distrofina.

15

Sumario

La presente invención se define por las reivindicaciones adjuntas. En particular, la presente invención proporciona;

20 [1]. Un sistema de direccionamiento de ADN que se une a un gen de gamma globina humana, comprendiendo el sistema de direccionamiento de ADN una proteína de fusión Cas y al menos un ARN guía (ARNg) que se dirige a una secuencia polinucleotídica con SEQ ID NO: 33, 34, 35 o 36, en donde el gen de gamma globina es HBG1 y/o HBG2.

25 [2]. El sistema de direccionamiento de ADN de [1], en donde la proteína de fusión Cas comprende dos dominios polipeptídicos heterólogos, en donde el primer dominio polipeptídico comprende una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (Cas) y el segundo dominio polipeptídico tiene una actividad seleccionada del grupo que consiste en actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa y actividad desmetilasa.

30 [3]. Un sistema de direccionamiento de ADN que se une a un gen de gamma globina humana, el sistema de direccionamiento de ADN que comprende al menos un ARN guía (ARNg) y una proteína de fusión que comprende dos dominios polipeptídicos heterólogos,

35 en donde el primer dominio polipeptídico comprende una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (Cas) y el segundo dominio polipeptídico tiene una actividad seleccionada del grupo que consiste en actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa y actividad desmetilasa,

40 en donde al menos un ARN guía (ARNg) se dirige a una secuencia polinucleotídica con SEQ ID NO: 33, 34, 35 o 36 y en donde el gen de gamma globina es HBG1 y/o HBG2.

[4]. El sistema de direccionamiento de ADN de [1] o [3], en donde el al menos un ARNg se dirige a una secuencia polinucleotídica con:

a) un complemento de SEQ ID NO: 33, 34, 35 o 36;

45 b) un ácido nucleico que es sustancialmente idéntico a SEQ ID NO: 33, 34, 35 o 36 o complemento de las mismas; o c) un ácido nucleico que se hibrida en condiciones estrictas con SEQ ID NO: 33, 34, 35 o 36, complemento de las mismas o una secuencia sustancialmente idéntica a las mismas.

50 [5]. El sistema de direccionamiento de ADN de [2] o [3], en donde la proteína Cas comprende una Cas9, en donde la Cas9 comprende al menos una mutación de aminoácido que anula la actividad nucleasa de Cas9.

[6]. El sistema de direccionamiento de ADN de [5], en donde la al menos una mutación de aminoácido es al menos una de D10A y H840A.

55 [7]. El sistema de direccionamiento de ADN de [6], en donde la proteína Cas comprende iCas9 (aminoácidos 36-1403 de SEQ ID NO: 1).

[8]. El sistema de direccionamiento de ADN de [2] o [3], en donde el segundo dominio polipeptídico de la proteína de fusión comprende al menos una repetición del dominio de activación de la transcripción VP16 o un dominio KRAB.

60 [9]. El sistema de direccionamiento de ADN de [8], en donde el segundo dominio polipeptídico de la proteína de fusión comprende un tetrámero de VP16 ("VP64") o un dominio de activación de p65.

[10]. El sistema de direccionamiento de ADN de [2] o [3], en donde la proteína de fusión comprende además un enlazador que conecta el primer dominio polipeptídico con el segundo dominio polipeptídico.

- [11]. El sistema de direccionamiento de ADN de [1] o [3], en donde el al menos un ARNg se dirige a una región promotora de HBG1 y/o HBG2.
- 5 [12]. Un polinucleótido aislado que codifica el sistema de direccionamiento de ADN de [1] o [3].
- [13]. Un vector que comprende el polinucleótido aislado de [12].
- [14]. Una célula que comprende el polinucleótido aislado de [12].
- 10 [15]. Un método para modular la expresión de genes de mamíferos en una célula, comprendiendo el método poner en contacto la célula con el sistema de direccionamiento de ADN de [1] o [3].
- [16]. Una composición para inducir la expresión génica en un sujeto, comprendiendo la composición el sistema de direccionamiento de ADN de [1] o [3].
- 15 [17]. Un vector lentivírico modificado que comprende un polinucleótido aislado que codifica el sistema de direccionamiento de ADN de [2] o [3], comprendiendo el polinucleótido aislado una primera secuencia polinucleotídica que codifica la proteína de fusión y una segunda secuencia polinucleotídica que codifica el al menos un ARNg.
- 20 [18]. El vector lentivírico modificado de [17] para su uso en un método de activación de la expresión génica endógena en un sujeto.
- [19]. El vector lentivírico modificado de [18] para el uso de [18], en donde el sujeto padece enfermedad de células falciformes o talasemia.
- 25 Adicionalmente, en el presente documento se divulga una proteína de fusión que comprende dos dominios polipeptídicos heterólogos. El primer dominio polipeptídico comprende una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (Cas) y el segundo dominio polipeptídico tiene una actividad seleccionada del grupo que consiste en actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa y actividad desmetilasa. La proteína Cas puede comprender Cas9. Cas9 puede comprender al menos una mutación de aminoácido que anula la actividad nucleasa de Cas9. La al menos una mutación de aminoácido puede ser al menos una de D10A y H840A. La proteína Cas puede comprender iCas9 (aminoácidos 36-1403 de SEQ ID NO: 1). El segundo dominio polipeptídico puede tener actividad de activación de la transcripción. El segundo dominio polipeptídico puede comprender al menos una repetición del dominio de activación de la transcripción de VP 16. El segundo dominio polipeptídico puede comprender un tetrámero de VP 16 ("VP64") o un dominio de activación de p65. La proteína de fusión puede comprender además un enlazador que conecta el primer dominio polipeptídico con el segundo dominio polipeptídico. La proteína de fusión puede comprender iCas9-VP64.
- 30 40 También se divulga en el presente documento un sistema de direccionamiento de ADN que comprende dicha proteína de fusión y al menos un ARN guía (ARNg). Al menos un ARNg puede comprender una secuencia polinucleotídica complementaria de 12-22 pares de bases de la secuencia de ADN diana seguida de un motivo adyacente al protoespaciador. El al menos un ARNg puede dirigirse a una región promotora de un gen, una región potenciadora de un gen o una región transcrita de un gen. El al menos un ARNg puede dirigirse a un intrón de un gen. El al menos un ARNg puede dirigirse a un exón de un gen. El al menos un ARNg puede dirigirse a la región promotora de un gen seleccionado del grupo que consiste en ASCL1, BRN2, MYT1L, NANOG, YEGFA, TERT, IL1B, IL1R2, IL1RN, HBG1, HBG2, y MYOD1. El al menos un ARNg puede comprender al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625.
- 45 50 También se divulga en el presente documento un sistema de direccionamiento de ADN que se une a un gen de la distrofina que comprende Cas9 y al menos un ARN guía (ARNg). El al menos un ARNg puede dirigirse a un intrón del gen de la distrofina. El al menos un ARNg puede dirigirse a un exón del gen de la distrofina. El al menos un ARN guía puede comprender al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625. El sistema de direccionamiento de ADN puede comprender entre uno y diez ARNg diferentes.
- 55 La presente divulgación se refiere a un polinucleótido aislado que codifica dicha proteína de fusión o dicho sistema de direccionamiento de ADN.
- 60 La presente divulgación se refiere a un vector que comprende dicho polinucleótido aislado.
- La presente divulgación se refiere a una célula que comprende dicho polinucleótido aislado o dicho vector.
- 65 La presente divulgación se refiere a un método para modular la expresión de genes de mamíferos en una célula. El método comprende poner en contacto la célula con dicha proteína de fusión, dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado o dicho vector. La expresión génica puede ser inducida.

- La presente divulgación se refiere a un método para transdiferenciar o inducir la diferenciación de una célula. El método comprende poner en contacto la célula con dicha proteína de fusión, dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado o dicho vector. La célula puede ser una célula de fibroblasto o una célula madre pluripotente inducida. La célula de fibroblasto puede transdiferenciarse en una célula neuronal o una célula miogena. El sistema de direccionamiento de ADN puede ponerse en contacto con la célula y al menos un ARNg se dirige a una región promotora de al menos un gen seleccionado del grupo que consiste en *ASCL1*, *BRN2*, *MYOD1* y *MYT1L*. El sistema de direccionamiento de ADN puede comprender al menos un ARNg que se dirige a la región promotora del gen *ASCL1* y al menos un ARNg que se dirige a la región promotora del gen *BRN2*. El sistema de direccionamiento de ADN puede comprender entre uno y veinte ARNg diferentes. El sistema de direccionamiento de ADN puede comprender 8 o 16 ARNg diferentes. El sistema de direccionamiento de ADN puede comprender dCas9-VP64. El sistema de direccionamiento de ADN puede suministrarse a la célula de forma vírica o no vírica.
- La presente divulgación se refiere a la corrección de un gen mutante en una célula. El método comprende administrar a una célula que contiene dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado o dicho vector. La corrección del gen mutante puede comprender una reparación dirigida por homología. El método puede comprender además administrar a la célula un ADN donante. El gen mutante puede comprender una mutación por desplazamiento del marco de lectura que provoca un codón de parada prematuro y un producto genético truncado. La corrección del gen mutante puede comprender la unión de extremos no homólogos mediada por nucleasa. La corrección del gen mutante puede comprender una eliminación de un codón de parada prematuro, una interrupción de un sitio aceptor de corte y empalme, una eliminación de uno o más exones o una interrupción de una secuencia donante de corte y empalme. La eliminación de uno o más exones puede dar lugar a la corrección del marco de lectura.
- La presente divulgación es para su uso en el tratamiento de un sujeto que lo necesita y que tiene un gen de distrofina mutante. El uso comprende administrar al sujeto dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado o dicho vector. El sujeto puede estar padeciendo distrofia muscular de Duchenne.
- La presente divulgación se refiere a la corrección de un gen de la distrofina mutante en una célula. El método comprende administrar a una célula que contiene un gen de distrofina mutante dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado, dicho vector o dicha célula. El gen de distrofina mutante puede comprender un codón de parada prematuro, un marco de lectura interrumpido a través de eliminación génica, un sitio aceptor de corte y empalme aberrante o un sitio donante de corte y empalme aberrante y en donde la región diana está en dirección en dirección 5' o en dirección 3' del codón de parada prematuro, el marco de lectura interrumpido, el sitio aceptor de corte y empalme aberrante o el sitio donante de corte y empalme aberrante. La corrección del gen de la distrofina mutante puede comprender una reparación dirigida por homología. El método puede comprender además administrar a la célula un ADN donante. El gen de la distrofina mutante puede comprender una mutación por desplazamiento del marco de lectura que provoca un codón de parada prematuro y un producto genético truncado. La corrección del gen de la distrofina mutante puede comprender la unión de extremos no homólogos mediada por nucleasa. La corrección del gen de la distrofina mutante puede comprender una eliminación de un codón de parada prematuro, una corrección de un marco de lectura alterado o una modulación del corte y el empalme mediante la alteración de un sitio aceptor de corte y empalme o la alteración de una secuencia donante de corte y empalme. La corrección del gen de la distrofina mutante puede comprender una eliminación de los exones 45-55 o del exón 51.
- La presente divulgación se refiere a un kit que comprende dicha proteína de fusión, dicho sistema de direccionamiento de ADN, dicho polinucleótido aislado, dicho vector o dicha célula.
- La presente divulgación se refiere a un método para modular la expresión de genes de mamíferos en una célula. El método comprende poner en contacto la célula con un polinucleótido que codifica un sistema de direccionamiento de ADN. El sistema de direccionamiento de ADN comprende dicha proteína de fusión y al menos un ARN guía (ARNg). El sistema de direccionamiento de ADN puede comprender entre uno y diez ARNg diferentes. Los diferentes ARNg pueden unirse a diferentes regiones diana dentro del gen diana. Las regiones diana pueden estar separadas por al menos un nucleótido. Las regiones diana pueden estar separadas por aproximadamente 15 a aproximadamente 700 pares de bases. Cada uno de los diferentes ARNg puede unirse a al menos un gen diana diferente. Los diferentes genes diana pueden estar ubicados en el mismo cromosoma. Los diferentes genes diana pueden estar ubicados en diferentes cromosomas. La al menos una región diana puede estar dentro de una región de cromatina no abierta, una región de cromatina abierta, una región promotora del gen diana, una región potenciadora del gen diana, una región transcrita del gen diana o una región en dirección 5' de un sitio de inicio de la transcripción del gen diana. La al menos una región diana puede estar ubicada entre aproximadamente 1 y aproximadamente 1000 pares de bases en dirección 5' de un sitio de inicio de la transcripción de un gen diana. La al menos una región diana puede estar ubicada entre aproximadamente 1 y aproximadamente 600 pares de bases en dirección 5' de un sitio de inicio de la transcripción de un gen diana. La expresión génica puede ser inducida. El sistema de direccionamiento de ADN puede comprender dos ARNg diferentes, tres ARNg diferentes, cuatro ARNg diferentes, cinco ARNg diferentes, seis ARNg diferentes, siete ARNg diferentes, ocho ARNg diferentes, nueve ARNg diferentes o diez ARNg diferentes. El al menos un ARN guía puede dirigirse a una región promotora de un gen seleccionado del grupo que consiste en *ASCL1*, *BRN2*, *MYT1L*, *NANOG*, *VEGFA*, *TERT*, *IL1B*, *IL1R2*, *IL1RN*, *HBG1*, *HBG2* y *MYOD1*. El al menos un ARN guía puede comprender al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625. La al menos una región diana puede estar

dentro de un intrón o un exón de un gen diana.

La presente divulgación se refiere a una composición para inducir la expresión de genes de mamíferos en una célula. La composición comprende dicha proteína de fusión y al menos un ARN guía (ARNg).

- 5 La presente divulgación se refiere a una composición para inducir la expresión de genes de mamíferos en una célula. La composición comprende una secuencia polinucleotídica aislada que codifica dicha proteína de fusión y al menos un ARN guía (ARNg). El al menos un ARN guía puede dirigirse a una región promotora de un gen seleccionado del grupo que consiste en *ASCL1*, *BRN2*, *MYT1L*, *NANOG*, *VEGFA*, *TERT*, *IL1B*, *IL1R2*, *IL1RN*, *HBG1*, *HBG2* y *MYOD1*.
- 10 El al menos un ARN guía puede comprender al menos una de SEQ ID NO 5-40, 65-144, 492-515, 540-563 y 585-625.
- La presente divulgación se refiere a una célula que comprende dicha composición para inducir la expresión de genes de mamíferos en una célula.
- 15 La presente divulgación se refiere a un kit que comprende dicha composición para inducir la expresión de genes de mamíferos en una célula o que comprende dicha célula dicha composición para inducir la expresión de genes de mamíferos en una célula.
- 20 La presente divulgación se refiere a un kit para inducir la expresión de genes de mamíferos en una célula. El kit comprende dicha composición para inducir la expresión de genes de mamíferos en una célula o que comprende dicha célula dicha composición para inducir la expresión de genes de mamíferos en una célula.
- 25 La presente divulgación se refiere a una composición para la edición genómica en un músculo de un sujeto. La composición comprende un vector de virus adenoasociado (AAV) modificado y una secuencia nucleotídica que codifica una nucleasa específica de sitio. El músculo es músculo esquelético o músculo cardíaco. El vector AAV modificado puede haber mejorado el tropismo del tejido del músculo cardíaco y esquelético. La nucleasa específica de sitio puede comprender una nucleasa con dedos de cinc, una nucleasa efectora TAL o un sistema CRISPR/Cas9. La nucleasa específica de sitio puede unirse a un gen o locus en la célula del músculo. El gen o locus puede ser el gen de la distrofina. La composición puede comprender además un ADN donante o un transgén.
- 30 La presente divulgación se refiere a un kit que comprende dicha composición para la edición genómica en un músculo de un sujeto.
- 35 La presente divulgación es para su uso en la edición genómica en un músculo de un sujeto. El uso comprende administrar al músculo dicha composición para la edición genómica en un músculo de un sujeto, en donde el músculo es músculo esquelético o músculo cardíaco. La edición genómica puede comprender corregir un gen mutante o insertar un transgén. La corrección de un gen mutante puede comprender eliminar, reorganizar o reemplazar el gen mutante. La corrección del gen mutante puede comprender la unión de extremos no homólogos mediada por nucleasa o la reparación dirigida por homología.
- 40 La presente divulgación es para su uso en el tratamiento de un sujeto. El uso comprende administrar dicha composición para la edición genómica en un músculo de un sujeto a un músculo del sujeto, en donde el músculo es músculo esquelético o músculo cardíaco. El sujeto puede estar padeciendo una afección del músculo esquelético o una enfermedad genética. El sujeto puede estar padeciendo distrofia muscular de Duchenne.
- 45 La presente divulgación se refiere a la corrección de un gen mutante en un sujeto, el uso comprende administrar dicha composición para la edición genómica en un músculo de un sujeto. El músculo es músculo esquelético o músculo cardíaco. La composición puede inyectarse en el músculo esquelético del sujeto. La composición puede inyectarse sistémicamente al sujeto. El músculo esquelético puede ser el músculo tibial anterior.
- 50 La presente divulgación se refiere a un vector lentivírico modificado para la edición genómica en un sujeto que comprende una primera secuencia polinucleotídica que codifica dicha proteína de fusión y una segunda secuencia polinucleotídica que codifica al menos un ARN_{gm}. La primera secuencia polinucleotídica puede estar unida operativamente a un primer promotor. El primer promotor puede ser un promotor constitutivo, un promotor inducible, un promotor reprimible o un promotor regulable. La segunda secuencia polinucleotídica puede codificar entre uno y diez ARN_{gm} diferentes. La segunda secuencia polinucleotídica puede codificar dos ARN_{gm} diferentes, tres ARN_{gm} diferentes, cuatro ARN_{gm} diferentes, cinco ARN_{gm} diferentes, seis ARN_{gm} diferentes, siete ARN_{gm} diferentes, ocho ARN_{gm} diferentes, nueve ARN_{gm} diferentes o diez ARN_{gm} diferentes. Cada una de las secuencias de polinucleótidos que codifican los diferentes ARN_{gm} pueden estar unidas operativamente a un promotor. Cada uno de los promotores unidos operativamente a los diferentes ARN_{gm} puede ser el mismo promotor. Cada uno de los promotores unidos operativamente a los diferentes ARN_{gm} pueden ser promotores diferentes. El promotor puede ser un promotor constitutivo, un promotor inducible, un promotor reprimible o un promotor regulable. El ARN_{gm} puede unirse a un gen diana. Cada uno de los ARN_{gm} puede unirse a una región diana diferente dentro de un loci diana. Cada uno de los ARN_{gm} puede unirse a una región diana diferente dentro de diferentes loci génicos. La proteína de fusión puede comprender la proteína Cas9 o la proteína iCas9-VP64. La proteína de fusión puede comprender un dominio VP64, un dominio ap300 o un dominio KRAB. Los dos o más genes endógenos pueden activarse transcripcionalmente. Los
- 65

dos o más genes endógenos pueden estar reprimidos.

- La presente divulgación se refiere a un método para activar un gen endógeno en una célula. El método comprende poner en contacto una célula con dicho vector lentivírico modificado. El gen endógeno puede activarse transitoriamente. El gen endógeno puede activarse de forma estable. El gen endógeno puede estar reprimido transitoriamente. El gen endógeno puede estar reprimido de forma estable. La proteína de fusión puede expresarse a niveles similares a los ARN_{gm}. La proteína de fusión se puede expresar a niveles diferentes a los ARN_{gm}. La célula puede ser una célula humana primaria.
- 10 La presente divulgación se refiere a un método de edición génica múltiple en una célula. El método comprende poner en contacto una célula con dicho vector lentivírico modificado. La edición génica múltiple puede comprender corregir al menos un gen mutante o insertar un transgén. La corrección de un gen mutante puede comprender eliminar, reordenar o reemplazar al menos un gen mutante. La corrección del al menos un gen mutante puede comprender la unión de extremos no homólogos mediada por nucleasa o la reparación dirigida por homología. La edición génica múltiple puede comprender eliminar al menos un gen, en donde el gen es un gen endógeno normal o un gen mutante.
- 15 15 La edición génica múltiple puede comprender eliminar al menos dos genes. La edición génica múltiple puede comprender eliminar entre dos y diez genes.
- 20 La presente divulgación se refiere a un método para modular la expresión génica de al menos un gen diana en una célula. El método comprende poner en contacto una célula con dicho vector lentivírico modificado. Puede modularse la expresión génica de al menos dos genes. Puede modularse la expresión génica de entre dos genes y diez genes. Puede modularse la expresión génica de al menos un gen diana cuando los niveles de expresión génica del al menos un gen diana aumentan o disminuyen en comparación con los niveles de expresión génica normales para el al menos un gen diana.
- 25 **Breve descripción de los dibujos**
- 30 La Fig. 1 muestra la activación guiada por ARN del gen *IL1RN* humano por iCas9-VP64. (a, b) Se creó un activador transcripcional guiado por ARN fusionando la Cas9 inactivada (iCas9, D10A/H840A) con el dominio de transactivación VP64. iCas9-VP64 reconoce sitios diana genómicos a través de la hibridación de un ARN guía (ARN_g) con una secuencia diana de 20 pb, (c) Plásmidos de expresión para cuatro ARN_g o ARN_{cr}/ARN_{ntracr} dirigidos a secuencias en el promotor de *IL1RN* se contráctan con el plásmido de expresión iCas9-VP64 en células HEK293T. La activación de la expresión de *IL1RN* se evaluó mediante qRT-PCR. (d) Los cuatro plásmidos de expresión de ARN_g se cotransfectaron con iCas9-VP64 individualmente o en combinación. Se observó una activación génica sólida mediante qRT-PCR solo en respuesta a la combinación de ARN_g. (e) La activación de la expresión de *IL1RN* se confirmó evaluando la secreción del producto génico IL-1ra en el medio mediante ELISA. IL-1ra solo se detectó en tres de las seis muestras tratadas con la combinación de ARN_g. Para (c-e), los datos se muestran como la media ± s.e.m. (n = 3 experimentos independientes). El tratamiento con la combinación de ARN_g fue estadísticamente diferente a todos los demás tratamientos (*P ≤ 0,02) según la prueba de Tukey, (f) Se realizó sec. de ARN en muestras tratadas con vector de expresión vacío (n = 2) o cotransfectadas con los plásmidos de expresión para iCas9-VP64 y los cuatro ARN_g dirigidos a *IL1RN* (n = 2). Los únicos cambios estadísticamente significativos en la expresión génica entre estos tratamientos fueron un aumento en las cuatro isoformas de *IL1RN* (tasa de descubrimiento falso ≤ 3 × 10⁻⁴) y una disminución en *IL32* (tasa de descubrimiento falso = 0,03).
- 35 45 La Fig. 2 muestra la activación guiada por ARN de genes humanos relevantes para la terapia celular y génica, la reprogramación genética y la medicina regenerativa. Se transfecaron células HEK293T con el plásmido de expresión iCas9-VP64 y cuatro ARN_g individualmente o en combinación. La expresión del gen diana se midió mediante qRT-PCR y se normalizó a niveles de ARNm de GAPDH. Los datos se muestran como la media ± s.e.m. (n = 3 experimentos independientes). El tratamiento con la combinación de ARN_g fue estadísticamente diferente a todos los demás tratamientos (*P < 0,05) según la prueba de Tukey.
- 50 55 La Fig. 3 muestra la expresión de iCas9-VP64. La expresión de iCas9-VP64 en células HEK293 transfecadas se confirmó mediante transferencia Western para la etiqueta del epítopo Flag N-terminal. El plásmido de expresión Cas9 ts no contiene la etiqueta del epítopo.
- 60 65 La Fig. 4 muestra las posiciones de los sitios diana de ARN_g y la hipersensibilidad a la DNase de genes diana humanos. Los cuatro sitios diana de ARN_g para cada locus se designan como pistas personalizadas encima de cada gen y los datos de sec. de DNase que indican regiones de cromatina abierta hipersensibles a la DNase se muestran debajo de cada gen. Se realizó sec. de DNase en células HEK293T para identificar regiones hipersensibles a DNase, como se describió anteriormente (Song *et al.*, *Cold Spring Harbor protocols* 2010, pdb prot5384 (2010); Song *et al.* *Genome Res* 21, 1757-1767 (2011)). Los resultados muestran que la cromatina abierta no era un requisito para la activación génica mediante combinaciones de ARN_g con iCas9-VP64.
- 65 La Fig. 5 muestra la ausencia de actividad nucleasa por parte de iCas9-VP64. Se cotransfectaron plásmidos de expresión Cas9 de tipo silvestre o iCas9-VP64 inactivados (D10A, H840A) con plásmidos de expresión para cuatro ARN guía diferentes dirigidos al promotor *IL1RN*. La actividad nucleasa se determinó mediante el ensayo de inspección

(Guschin *et al.*, *Methods Mol Biol* 649, 247-256 (2010)). Las bandas de menor peso molecular indicativas de actividad nucleasa y reparación del ADN mediante unión de extremos no homólogos solo están presentes después del tratamiento con Cas9 de tipo silvestre, apoyando la anulación de la actividad nucleasa por iCas9-VP64.

- 5 La Fig. 6 muestra sec. de ARN para muestras tratadas con ARNg dirigidos a *HBG1* y *HBG2*. Se realizó sec. de ARN en muestras tratadas con un vector de expresión vacío ($n = 3$) o cotransfectadas con los plásmidos de expresión para iCas9-VP64 y los cuatro ARNg dirigidos a *HBG1* ($n = 2$). Tres de estos ARNg también se dirigen a *HBG2*. Se observaron aumentos en ambos *HBG1* y *HBG2* con respecto al control pero no fueron estadísticamente significativos debido a los bajos niveles de expresión. Los únicos cambios estadísticamente significativos en la expresión génica entre estos tratamientos fueron disminuciones en *IL32* (tasa de descubrimiento falso = 0,0007) y *TNFRS9* (tasa de descubrimiento falso = 0,002).
- 10

La Fig. 7 muestra la regulación positiva de *Ascl1* y γ -globina por iCas9-VP64. Las células HEK293T se transfecaron con iCas9-VP64 y cuatro ARNg dirigidos al promotor de *ASCL1* o de *HBG1*. Los niveles de producción de proteínas *Ascl1* y γ -globina correspondientes se evaluaron mediante transferencia Western. Se detectaron niveles bajos de estas proteínas en células HEK293T y se detectaron aumentos en la expresión después del tratamiento con iCas9-VP64 en dos experimentos independientes.

- 20 La Fig. 8 muestra la activación de dianas en dirección 3' de *Ascl1* en fibroblastos embrionarios murinos tratados con iCas9-VP64. Se transfecaron fibroblastos embrionarios de ratón (Mouse Embryonic Fibroblasts, MEF) con un plásmido de expresión de GFP de control o el plásmido de expresión iCas9-VP64 y una combinación de cuatro plásmidos de expresión de ARNg dirigidos a *ASCL1* en una proporción de 50:50 o 75:25. (a) Los sitios diana del ARNg en el promotor de *ASCL1* humano (SEQ ID NO: 3) se conservan en el promotor de *ASCL1* de ratón (SEQ ID NO: 4). Los sitios diana se indican con líneas continuas y la región transcrita se indica con una línea discontinua, (b) La expresión de *ASCL1* en MEF aumentó en dos días después del tratamiento con iCas9-VP64/ARNg según lo determinado por qRT-PCR. (c-h) Despues de 10 días en medio de inducción neuronal, las células se tiñeron para *Ascl1* y *Tuj1*, un marcador temprano de diferenciación neuronal (c-d) o para *Tuj1* y *MAP2*, un marcador de diferenciación neuronal más madura (d-f). Algunas células positivas para *Tuj1* adoptaron morfologías neuronales (f-g) y se encontró que una sola célula era positiva para *Tuj1* y *MAP2* (g). (h) Las células positivas para *Tuj1* se identificaron fácilmente en los cultivos tratados con iCas9-VP64/ARNg (~0,05 %) pero estuvieron ausentes en los controles, $n = 3$ muestras independientes y los datos se representan como media \pm error típico de la media. El ARNg 75/25 es significativamente diferente del ARNg 50/50 y del control (* $P < 0,01$, prueba de Tukey).
- 30

- 35 La Fig. 9 muestra (a) la secuencia de la proteína iCas9-VP64 (SEQ ID NO: 1) y (b) la secuencia del casete de expresión de ARNg con el promotor U6 (SEQ ID NO: 2).

La Fig. 10 muestra las curvas patrón para qRT-PCR. Para cada gen, la muestra experimental con el nivel de expresión más alto se diluyó para crear una curva patrón que se analizó mediante qRT-PCR para garantizar una amplificación eficiente en un intervalo dinámico apropiado. Las eficiencias de todas las reacciones de amplificación estuvieron dentro del 90-115 %.

- 40 Las Fig. 11(a)-11(b) muestran la validación de la reparación guiada por ARN. La Fig. 11(a) muestra los resultados del ensayo de inspección del ADN genómico recogido de células HER 293T dos días después de que Cas9 se cotransfectara en las células con un vector vacío (control negativo) o ARNg. La Fig. 11(b) muestra la ubicación de la diana de ARNg. La Fig. 11(c) muestra los tamaños de escisión esperados para cada ARNg.

45 La Fig. 12 muestra la reparación guiada por ARN en células DMD 8036 (del48-50) como se muestra mediante el ensayo de inspección.

- 50 La Fig. 13 muestra la reparación guiada por ARN en células DMD 8036 (del48-50) como se muestra mediante PCR en todo el locus. La PCR de un gen de distrofina de tipo silvestre genera un fragmento de 1447 pb de tamaño, mientras que la PCR del gen mutante en la línea celular DMD 8036 muestra una eliminación de aproximadamente 817 pb. La banda de eliminación después de la introducción del sistema basado en CRISPR/Cas9 fue de aproximadamente 630 pb.
- 55

La Fig. 14 muestra la reparación guiada por ARN en células DMD 8036 (del48-50) como se muestra mediante transferencia Western con MANDYS8 (anticuerpo anti-distrofina) y anticuerpo GAPDH (control positivo).

- 60 La Fig. 15 muestra datos de secuenciación de ChIP que ilustran la unión específica de iCas9-VP64 dirigida al promotor IL1RN. Las células HER 293T se transfecaron con iCas9-VP64 dirigida al promotor IL1RN.

65 La Fig. 16 muestra CRISPR^zCas9 dirigido al gen de la distrofina. (A) las secuencias de ARNgm se diseñaron para unirse a secuencias en la región del punto caliente mutacional del exón 45-55 del gen de la distrofina, de tal manera que la edición génica podría restaurar la expresión de distrofina a partir de una amplia diversidad de mutaciones específicas del paciente. Las flechas dentro de los intrones indican dianas de ARNgm diseñadas para eliminar exones completos del genoma. Las flechas dentro de los exones indican dianas de ARNgm diseñadas para crear

- desplazamientos del marco específicos en el gen de la distrofina. (B) Ejemplo de corrección del marco después de la introducción de pequeñas inserciones o eliminaciones mediante la reparación del ADN NHEJ en el exón 51 usando el ARN_{gm} CR3. (C) Esquema de dianas de ARN_{gm} múltiplex diseñadas para eliminar el exón 51 y restaurar el marco de lectura de la distrofina en una mutación del paciente con la eliminación de los exones 48-50. (D) Esquema de dianas de ARN_{gm} múltiplex diseñadas para eliminar la región entera del exón 45-55 para abordar una diversidad de mutaciones de pacientes con DMD.
- La Fig. 17 muestra imágenes de geles TBE-PAGE usados para cuantificar los resultados del ensayo de inspección para medir la modificación génica del día 3 en la Tabla 7. Los asteriscos marcan los tamaños esperados de bandas indicativas de actividad nucleasa.
- La Fig. 18 muestra imágenes de geles TBE-PAGE usados para cuantificar los resultados del ensayo de inspección para medir la modificación génica del día 10 en la Tabla 7. Los asteriscos marcan los tamaños esperados de bandas indicativas de actividad nucleasa.
- La Fig. 19 muestra la clasificación de flujo activada por fluorescencia para enriquecer mioblastos DMD modificados genéticamente. (A) Un plásmido que expresa una proteína SpCas9 optimizada con codones humanos unida a un marcador de GFP usando una secuencia peptídica de omisión ribosómica T2A se coelectroporó en mioblastos DMD humanos con uno o dos plásmidos que portaban casetes de expresión de ARN_{gm}. (B) Los casetes de expresión de ARN_{gm} indicados se cotransfectaron de forma independiente en HEK293T con un plásmido separado que expresa SpCas9 con (abajo) o sin (arriba) un marcador de GFP enlazado a SpCas9 mediante una secuencia peptídica de omisión ribosómica T2A. Las frecuencias de modificación génica se evaluaron 3 días después de la transfección mediante el ensayo de inspección. (C) Los mioblastos de DMD con eliminaciones de los exones 48-50 en el gen de la distrofina se trataron con ARN_{gm} que corrigen el marco de lectura de la distrofina en estas células de pacientes. La modificación génica se evaluó 20 días después de la electroporación en células sin clasificar (a granel) o clasificadas GFP+. (D) Expresión de GFP en mioblastos DMD 3 días después de la electroporación con plásmidos de expresión indicados. Las eficiencias de transfección y las poblaciones de células clasificadas están indicadas por la región clasificada.
- La Fig. 20 muestra desplazamientos de marco específicos para restaurar el marco de lectura de distrofina usando CRISPR/Cas9. (A) La región 5' del exón 51 se marcó como diana usando un ARN_{gm}, CR3, que se une inmediatamente en dirección 5' del primer codón de parada fuera del marco. PAM: motivo adyacente al protoespaciador. (B) El locus del exón 51 se amplificó por PCR a partir de células HEK293T tratadas con casetes de expresión SpCas9 y CR3. Las secuencias de clones individuales se determinaron mediante secuenciación de Sanger. La secuencia superior (en negrita, exón en rojo) es la secuencia nativa no modificada. El número de clones para cada secuencia se indica entre paréntesis. (C) Resumen de la eficiencia total de edición génica y las conversiones del marco de lectura resultantes de la modificación génica que se muestra en (B). (D) Transferencia Western para la expresión de distrofina en mioblastos de DMD humanos tratados con SpCas9 y el casete de expresión de ARN_{gm} CR3 (Fig. 19C) para crear desplazamientos de marco específicos para restaurar el marco de lectura de la distrofina. La expresión de distrofina se sondó usando un anticuerpo contra el dominio de varilla de la proteína distrofina después de 6 días de diferenciación.
- La Fig. 21 muestra la eliminación del exón 51 del genoma humano usando edición del gen CRISPR/Cas9 multiplex. (A) PCR genómica de punto final a través del locus del exón 51 en mioblastos DMD humanos con una eliminación de los exones 48-50. La flecha superior indica la posición esperada de los amplicones de PCR de longitud completa y las dos flechas inferiores indican la posición esperada de los amplicones de PCR con eliminaciones provocadas por las combinaciones de ARN_{gm} indicadas. (B) Se clonaron productos de PCR de (A) y se secuenciaron clones individuales para determinar las inserciones y eliminaciones presentes en el locus diana. La fila superior muestra la secuencia no modificada de tipo silvestre y los triángulos indican los sitios de escisión de SpCas9. A la derecha hay cromatogramas representativos que muestran las secuencias de las uniones de eliminación esperadas. (C) Análisis de RT-PCR de punto final de transcripciones de ARNm de distrofina en mioblastos DMD Δ48-50 humanos modificados con CRISPR/Cas9 tratados con los ARN_{gm} indicados. A la derecha se muestra un cromatograma representativo del producto de PCR de eliminación esperado. Asterisco: banda resultante de la hibridación de la cadena del producto de eliminación con la cadena no modificada. (D) El rescate de la expresión de la proteína distrofina mediante la edición genómica CRISPR/Cas9 se evaluó mediante transferencia Western para la proteína distrofina con GAPDH como un control de carga. La flecha indica la banda de distrofina restaurada esperada.
- La Fig. 22 muestra la eliminación de toda la región del exón 45-55 en mioblastos DMD humanos mediante edición múltiple del gen CRISPR/Cas9. (A) PCR genómica de punto final de ADN genómico para detectar la eliminación de la región entre el intrón 44 y el intrón 55 después de tratar mioblastos HEK293T o DMD con los ARN_{gm} indicados. (B) Se analizaron clones individuales de productos de PCR del tamaño esperado para las eliminaciones de mioblastos DMD en (A) mediante secuenciación de Sanger para determinar las secuencias de eliminaciones genómicas presentes en el locus diana. A continuación se muestran cromatogramas representativos que muestran la secuencia de las uniones de eliminación esperadas. (C) Análisis de RT-PCR de punto final de transcripciones de ARNm de distrofina en mioblastos DMD Δ48-50 humanos modificados con CRISPR/Cas9 tratados con los ARN_{gm} indicados. A la derecha se muestra un cromatograma representativo del producto de PCR de eliminación esperado. (D) Análisis de la expresión

de la proteína distrofina restaurada mediante transferencia Western después de la electroporación de mioblastos DMD con ARN_{gm} dirigidos al intrón 44 y/o al intrón 55.

- 5 La Fig. 23 muestra la verificación del enriquecimiento basado en citometría de flujo de mioblastos DMD modificados genéticamente usados para experimento de trasplante de células *in vivo*. Los mioblastos DMD se trataron con Cas9 con o sin vectores de expresión de ARN_{gm} para CR1 y CR5 y se clasificaron para células GFP+ mediante citometría de flujo. Las eliminaciones en el locus del exón 51 se detectaron mediante PCR de punto final usando cebadores que flanquean el locus. Ctrl neg: Mioblastos DMD tratados solo con Cas9 y clasificados para células GFP+.
- 10 La Fig. 24 muestra la expresión de distrofina humana restaurada *in vivo* después del trasplante de mioblastos de DMD humanos tratados con CRISPR/Cas9 en ratones inmunodeficientes. Se trataron mioblastos humanos Δ48-50 DMD con SpCas9, CR1 y CR5 para eliminar el exón 51 y se clasificaron para la expresión de GFP como se muestra en la Fig. 19. Estas células clasificadas y las células de control no tratadas se inyectaron en las extremidades traseras de ratones inmunodeficientes y se evaluó la expresión de proteínas humanas específicas en fibras musculares después de 4 semanas después del trasplante. Las criosecciones se tiñeron con anti-espectrina humana, que se expresa mediante mioblastos corregidos y no corregidos que se han fusionado en miofibra de ratón, o anticuerpos anti-distrofina humana, según se indique. Las flechas blancas indican fibras musculares positivas para distrofina humana.
- 15 La Fig. 25 muestra imágenes de inmunofluorescencia adicionales que prueban la expresión de distrofina humana. En la parte superior izquierda se muestran secciones seriadas de regiones teñidas con anti-espectrina humana. (A-C) Secciones de músculos inyectados con mioblastos DMD humanos no tratados. (D-F) Secciones de músculos inyectados con mioblastos DMD humanos tratados con CR1/5 enriquecidos por citometría de flujo. Las flechas blancas indican fibras positivas para distrofina.
- 20 La Fig. 26 muestra la evaluación de la toxicidad de CRISPR/Cas9 y los efectos inespecíficos para la eliminación del exón 51 mediada por CR1/CR5 en células humanas. (A) Resultados de un ensayo de citotoxicidad en células HEK293T tratadas con SpCas9 optimizado para humanos y las construcciones de ARN_{gm} indicadas. La citotoxicidad se basa en la supervivencia de las células positivas para GFP que se cotransfectan con la nucleasa indicada. I-SceI es una meganucleasa no tóxica bien caracterizada y GZF3 es una nucleasa de dedos de cinc tóxica conocida. (B) Análisis de inspección en sitios inespecíficos en células hDMD clasificadas tratadas con casetes de expresión que codifican Cas9 para los ARN_{gm} indicados. Estos tres sitios inespecíficos probados en células hDMD se identificaron a partir de un panel de 50 sitios predichos probados en células HEK293T (Fig. 27 y Tabla 4). TGT: locus diana para el ARN_{gm} indicado. OT: locus inespecífico. (C, D) PCR anidada de punto final para detectar translocaciones cromosómicas en (C) células HEK293T tratadas con Cas9 y CR1 o (D) células hDMD clasificadas tratadas con Cas9, CR1 y CR5. El esquema muestra la ubicación relativa de los pares de cebadores anidados personalizados para cada evento de translocación. El tamaño esperado de cada banda se estimó en función del tamaño del cebador y la ubicación del sitio de corte de ARN_{gm} previsto en cada locus. Los asteriscos indican bandas detectadas en el tamaño esperado. Las identidades de las bandas en (C) se verificaron mediante secuenciación de Sanger desde cada extremo (Fig. 30). Se muestra un cromatograma representativo de la translocación de P2/P5 en células HEK293T.
- 25 La Fig. 27 muestra imágenes de geles TBE-PAGE usados para cuantificar los resultados del ensayo de inspección para medir la modificación génica diana e inespecífica en la Tabla 4. Los asteriscos marcan los tamaños esperados de bandas indicativas de actividad nucleasa.
- 30 La Fig. 28 muestra una PCR anidada de punto final para detectar translocaciones cromosómicas provocadas por actividad inespecífica de CRISPR/Cas9 para CR3 y CR6/CR36 en células humanas. Se usó un análisis de PCR de punto final anidada para detectar translocaciones en (A) HEK293T o células hDMD clasificadas tratadas con Cas9 y CR3 como se indica, (B) células HEK293T tratadas con Cas9 y CR36 solo o (C) células hDMD clasificadas tratadas con casetes de expresión Cas9, CR6 y CR36. La segunda reacción de PCR anidada para la translocación se amplificó usando cebadores personalizados para cada locus de translocación previsto para maximizar la especificidad (véase la Tabla 4). El esquema representa la ubicación relativa de los pares de cebadores anidados usados para detectar la presencia de translocaciones. Cada posible evento de translocación se amplificó primero a partir de ADN genómico aislado de células tratadas con o sin los ARN_{gm} indicados. Se realizó una segunda reacción de PCR anidada usando cebadores dentro de los amplicones de PCR previstos que resultarían de las translocaciones. El tamaño esperado se estimó en función del sitio de unión del cebador indicado y el sitio de corte de ARN_{gm} previsto en cada locus. * indica bandas detectadas en el tamaño esperado y verificadas mediante secuenciación Sanger desde cada extremo. Indica amplicones en los que la secuenciación de Sanger mostró secuencias distintas a la translocación predicha, probablemente sea el resultado de un cebado incorrecto durante la PCR anidada.
- 35 La Fig. 29 muestra chromatogramas de secuenciación de Sanger para las bandas detectadas en la Fig. 28 resultantes de translocaciones entre CR3 y CR3-OT1, en los cromosomas X y 1, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR3. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.
- 40 La Fig. 30 muestra imágenes de geles TBE-PAGE usados para cuantificar los resultados del ensayo de inspección para medir la modificación génica diana e inespecífica en la Tabla 4. Los asteriscos marcan los tamaños esperados de bandas indicativas de actividad nucleasa.
- 45 La Fig. 31 muestra una PCR anidada de punto final para detectar translocaciones cromosómicas provocadas por actividad inespecífica de CRISPR/Cas9 para CR3 y CR6/CR36 en células humanas. Se usó un análisis de PCR de punto final anidada para detectar translocaciones en (A) HEK293T o células hDMD clasificadas tratadas con Cas9 y CR3 como se indica, (B) células HEK293T tratadas con Cas9 y CR36 solo o (C) células hDMD clasificadas tratadas con casetes de expresión Cas9, CR6 y CR36. La segunda reacción de PCR anidada para la translocación se amplificó usando cebadores personalizados para cada locus de translocación previsto para maximizar la especificidad (véase la Tabla 4). El esquema representa la ubicación relativa de los pares de cebadores anidados usados para detectar la presencia de translocaciones. Cada posible evento de translocación se amplificó primero a partir de ADN genómico aislado de células tratadas con o sin los ARN_{gm} indicados. Se realizó una segunda reacción de PCR anidada usando cebadores dentro de los amplicones de PCR previstos que resultarían de las translocaciones. El tamaño esperado se estimó en función del sitio de unión del cebador indicado y el sitio de corte de ARN_{gm} previsto en cada locus. * indica bandas detectadas en el tamaño esperado y verificadas mediante secuenciación Sanger desde cada extremo. Indica amplicones en los que la secuenciación de Sanger mostró secuencias distintas a la translocación predicha, probablemente sea el resultado de un cebado incorrecto durante la PCR anidada.
- 50 La Fig. 32 muestra chromatogramas de secuenciación de Sanger para las bandas detectadas en la Fig. 31 resultantes de translocaciones entre CR3 y CR3-OT1, en los cromosomas X y 1, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR3. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.
- 55 La Fig. 33 muestra chromatogramas de secuenciación de Sanger para las bandas detectadas en la Fig. 32 resultantes de translocaciones entre CR3 y CR3-OT1, en los cromosomas X y 1, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR3. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.
- 60 La Fig. 34 muestra chromatogramas de secuenciación de Sanger para las bandas detectadas en la Fig. 33 resultantes de translocaciones entre CR3 y CR3-OT1, en los cromosomas X y 1, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR3. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.
- 65 La Fig. 35 muestra chromatogramas de secuenciación de Sanger para las bandas detectadas en la Fig. 34 resultantes de translocaciones entre CR3 y CR3-OT1, en los cromosomas X y 1, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR3. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.

La Fig. 30 muestra chromatogramas de secuenciación de Sanger para bandas detectadas en la Fig. 26C resultantes de translocaciones entre CRT y CR1-OT1, en los cromosomas X y 16, respectivamente, en células HEK293T tratadas con casetes de genes Cas9 y CR1. Las flechas muestran regiones de homología con el cromosoma indicado cerca de los puntos de ruptura esperados provocados por los ARN_{gm} apropiados. Obsérvese que las lecturas de secuenciación se desfasan cerca del punto de ruptura debido a la naturaleza propensa a errores de la reparación del ADN mediante uniones de extremos no homólogos.

- 5 La Fig. 31 muestra una descripción general de inyecciones de AAV *in vivo* y recolección de tejido.
- 10 La Fig. 32 muestra el análisis de inspección de las actividades de Rosa26 ZFN en músculo esquelético *in vitro* e *in vivo* después del suministro de AAV-SASTG-ROSA. Las flechas indican las bandas esperadas resultantes de la escisión de inspección, n.d.: no detectado, (a) C2C12 en proliferación se transdujeron con la cantidad indicada de virus y se recogieron 4 días después de la infección. Las flechas indican los tamaños de banda esperados resultantes de la escisión de inspección, (b) C2C12 se incubaron en medio de diferenciación durante 5 días y después se transdujeron 15 con la cantidad indicada de virus AAV-SASTG-ROSA en placas de 24 pocillos. Las muestras se recogieron 10 días después de la transducción, (c) La cantidad indicada de AAV-SASTG-ROSA se injectó directamente en el tibial anterior de ratones C57BL/6J y los músculos se recogieron 4 semanas después de la infección. Los músculos TA recogidos se dividieron en 8 trozos separados para el análisis de ADN genómico, cada uno mostrado en un carril separado.
- 20 La Fig. 33 muestra la secuencia de ADN opt de Rosa T2A (SEQ ID NO: 434) y la secuencia de proteína opt de Rosa T2A (SEQ ID NO: 435).
- 25 La Fig. 34 muestra la secuencia de ADN de la cápside de SASTG (SEQ ID NO: 436) y la secuencia del péptido de la cápside de SASTG (SEQ ID NO: 437).
- 30 La Fig. 35 muestra la secuencia del sitio diana DZF16 ZFN (SEQ ID NO: 442), la secuencia de aminoácidos completa izquierda de DZF16-L6 (SEQ ID NO: 443) y la secuencia de aminoácidos completa derecha de DZF16-R6 (SEQ ID NO: 444).
- 35 La Fig. 36 muestra la secuencia del sitio diana E51C3 ZFN (SEQ ID NO: 445), la secuencia de aminoácidos completa izquierda de E51C3-3L (SEQ ID NO: 446) y la secuencia de aminoácidos completa derecha de E51C3-3R (SEQ ID NO: 447).
- 40 La Fig. 37 muestra la secuencia del sitio diana DZF15 ZFN (SEQ ID NO: 448), la secuencia de aminoácidos completa izquierda de DZF15-L6 (SEQ ID NO: 449), la secuencia de aminoácidos completa derecha de DZF15-R6 (SEQ ID NO: 450), la secuencia de aminoácidos completa izquierda de DZF15-L5 (SEQ ID NO: 451), la secuencia de aminoácidos completa derecha de DZF15-R5 (SEQ ID NO: 452).
- 45 La Fig. 38 muestra la secuencia del sitio diana E51C4 ZFN (SEQ ID NO: 453), la secuencia de aminoácidos completa izquierda de E51C4-4L (SEQ ID NO: 454) y la secuencia de aminoácidos completa derecha de E51C4-4R (SEQ ID NO: 455).
- 50 La Fig. 39 muestra diagramas esquemáticos de un "Vector individual, sistema CRISPR múltiple", "Vector doble, sistema CRISPR múltiple", y "Vector individual, sistema de ARN_g individual".
- 55 La Fig. 40 muestra las secuencias de nucleótidos de SaCas9-NLS (con NLS subrayada) (SEQ ID NO: 64) y ARN_g de SaCas9 (SEQ ID NO: 116).
- 60 La Fig. 41 muestra las secuencias de nucleótidos de NmCas9 (con NLS 1 subrayado, NLS 2 subrayado y en negrita y la etiqueta HA en negrita), Horquilla corta de NmCas9 de Thomson PNAS 2013 (SEQ ID NO: 118) y horquilla larga de NmCas9 de Church Nature Biotech 2013 (SEQ ID NO: 119).
- 65 La Fig. 42 muestra la validación de construcciones de expresión de ARN_{gm} y Cas9 lentivírica, (a) Se transfecionaron construcciones que codifican promotores Pol III únicos que expresan ARN_{gm} dirigidos al locus AAVS1 o una construcción que contiene el promotor hU6 inmediatamente seguido de politimidina para terminar la expresión ("PoliT") en células HEK293T. Se usó RT-PCR de punto final para sondear la expresión de cada construcción de promotor/ARN_{gm} indicada dos días después de la transfección. -RT: sin control de transcriptasa inversa, (b) Las HEK293T se transfecionaron con vectores de expresión que codifican la nucleasa de dedos de cinc AAVS1 o Cas9-T2A-GFP y los casetes de expresión de promotor/ARN_{gm} indicados y se evaluaron los niveles de modificación génica 3 días después de la transfección usando el ensayo de inspección, (c) Se transdujeron células HEK293T con construcciones lentivíricas que codifican las construcciones Cas9-T2A-GFP indicadas sin ARN_{gm} y se evaluó la expresión de Cas9 mediante transferencia Western 7 días después de la transducción mediante sonda para detectar una etiqueta de epítopo FLAG en el extremo N de la proteína Cas9.
- 66 La Fig. 43 muestra el ensamblaje Golden Gate de casetes de expresión CRISPR/Cas9 lentivíricos individuales.

La figura 44 muestra el suministro lentivírico individual de un sistema CRISPR/Cas9 múltiplex, (a) Se clonaron cuatro ARNgm dirigidos a distintos loci genómicos en un vector lentivírico que expresa la nucleasa Cas9 activa, (b) Se transdijeron HEK293T y fibroblastos dérmicos humanos primarios con lentivirus que expresaban los ARNgm indicados y se analizaron los eventos de escisión usando el ensayo de inspección. Las HEK293T se analizaron 7 días después de la transducción. Los fibroblastos humanos se analizaron 10 días después de la transducción.

La Fig. 45 muestra la activación génica transitoria en HEK293T que expresan de manera estable dCas9-VP64. Las HEK293T se transdijeron con lentivirus para expresar de manera estable dCas9-VP64 y posteriormente se transfecaron con plásmido que expresaba las combinaciones de ARNgm indicadas. Al variar la cantidad de ARNgm suministrados, la activación génica endógena sintonizable de los loci endógenos IL1RN (a) y HBG1 (b) se logró 3 días después de la transfección. Se observaron niveles máximos de IL1RN (c) y HBG1 (d) endógenos 3-6 días después de la transfección y el nivel de activación volvió a los niveles de fondo entre los días 15-20. De manera importante, las líneas celulares pudieron reaccionar después de una segunda transfección el día 20 aunque a un nivel más bajo que el observado anteriormente.

La Fig. 46 muestra la activación génica estable en HEK293T usando un único vector lentivírico multiplex dCas9-VP64. Las HEK293T se transdijeron con lentivirus para expresar de manera estable dCas9-VP64 y las combinaciones indicadas de ARNg. Al variar la cantidad de ARNgm suministrados, la activación génica endógena sintonizable de los loci endógenos IL1RN (a) y HBG1 (b) se logró 7 días después de la transducción. Se observaron niveles máximos de IL1RN (c) y HBG1 (d) endógenos 6 días después de la transducción y el nivel de activación se mantuvo hasta el día 21.

La Fig. 47 muestra los niveles de expresión de ARNm de IL1RN.

La Fig. 48 muestra un esquema que representa la conversión directa de fibroblastos en neuronas mediante la expresión ectópica de los factores de transcripción neuronal BAM.

La Fig. 49 muestra (A) Esquema de la construcción dCas9-VP64. dCas9-VP64 es una forma catalíticamente inactiva de la proteína Cas9 fusionada a un tetramero del dominio de activación transcripcional VP 16. (B) Esquema que muestra el mecanismo de reclutamiento guiado por ARN de dCas9-VP64 a una diana genómica. (C) Esquema del protocolo experimental para generar iN con factores de transcripción CRISPR/Cas9.

La Fig. 50 muestra la expresión de ASCL1 endógena el día 3 determinada por (A) qRT-PCR o la proteína ASCL1 total detectada por (B) inmunofluorescencia en MEF transducidos con dCas9-VP64 y transfecidos con ARNg dirigidos al promotor de ASCL1, ADNc de ASCL1 o luciferasa. El asterisco (*) indica un aumento significativo ($p<0,05$) en la expresión de ASCL1 con el cosuministro de 8 ARNg en comparación con 4 ARNg. La expresión ectópica de ASCL1 produjo más proteína que la inducida por dCas9-VP64 y 8 ARNg dirigidos al promotor Ascl1, pero no activó el locus endógeno en el día 3 en cultivo.

La figura 51 muestra (A) células positivas para TUJ1 y MAP2 generadas por factores BAM ectópicos o por dCas9-VP64 y ARNg dirigidos a los promotores de BRN2 y de ASCL1 (B) Células con morfología neuronal que expresan un indicador hSyn-RFP el día 11 en medio N3.

La Fig. 52 muestra (A) Una célula con morfología neuronal positiva para el indicador de calcio GCaMP5 en presencia (abajo) o ausencia (arriba) de KC1 en el medio de cultivo. (B) Un rastro de intensidad fluorescente normalizada a lo largo del tiempo que muestra la despolarización de la célula en respuesta a la adición de KC1.

La Fig. 52 muestra la activación de dianas posteriores de Ascl1 y Brn2, es decir, genes reguladores maestros, en fibroblastos embrionarios murinos tratados con iCas9-VP64 usando factores de transcripción dCas9-VP64 para convertir los fibroblastos en neuronas. Se transfecaron fibroblastos embrionarios de ratón (MEF) con un plásmido de expresión de GFP de control o el plásmido de expresión iCas9-VP64 y una combinación de ocho plásmidos de expresión de ARNg dirigidos a ASCL1 y BRN2. Los factores de transcripción dCas9 se suministraron de forma vírica. Después de 10 días en medio de inducción neuronal, las células se tiñeron para Tuj1, un marcador temprano de diferenciación neuronal y MAP2, un marcador de diferenciación neuronal más madura. La conversión a neuronas fue eficiente.

La Fig. 53 muestra la plataforma CRISPR/Cas9 para el control de la regulación génica de mamíferos. A. Los efectores basados en Cas9 se unen a secuencias genómicas en presencia de una molécula de ARNg químérica que consiste en una región constante que forma complejos con Cas9 precedida por un protoespaciador intercambiable de 20 pb que confiere especificidad del sitio diana. B. Los factores de transcripción sintéticos basados en Cas9 reprimen la transcripción de un gen diana al interferir con la actividad de la ARN polimerasa o al unirse dentro del promotor y bloquear los sitios de unión de los factores de transcripción endógenos. C. Dirigirse a elementos reguladores tales como los potenciadores también podría bloquear potencialmente la expresión de múltiples genes distales.

La Fig. 54 muestra cómo dirigirse al potenciador HS2 usando CRISPR/dCas9-KRAB. La región HS2 es un potente

potenciador que regula distalmente la expresión de genes de globina >10 kb en dirección 3'. Se diseñó un panel de ARNg individuales para dirigirse a sitios a lo largo de la región potenciadora.

5 La Fig. 55 muestra que los ARNg individuales dirigidos al potenciador HS2 efectúan una potente represión transcripcional de los genes de globina. A. Los represores dCas9 y dCas9-KRAB se suministraron en un vector lentivírico. Se transfecaron transitoriamente ARNg individuales para su detección. Cuando se analiza mediante RT-PCR cuantitativa 3 días después de la transfección, Los K562 que expresan dCas9-KRAB logran hasta un 80 % de represión de los genes de B. γ -globina, C. ϵ -globina y D. β -globina, en comparación con las células control que no recibieron tratamiento con ARNg. D. La expresión de proteínas en células que expresan dCas9 o dCas9-KRAB y tratadas con Cr4 o Cr8 muestra una leve represión de la expresión de γ -globina el día 3, en comparación con los controles de β -actina.

10 15 La Fig. 56 muestra la expresión de genes del locus de globina con dosis variables de plásmido de ARNg suministrado a células tratadas con A. sin lentivirus, B. lentivirus dCas9 o C. lentivirus dCas9-KRAB. El aumento de la dosis del plásmido Cr4 ARNg produjo una represión mejorada en las células tratadas con dCas9-KRAB, lo que indica que tanto el efecto dCas9-KRAB como el ARNg específico desempeñan un papel en el logro de la represión.

20 25 La Fig. 57 muestra que el suministro estable de ARNg individuales con dCas9-KRAB silencia la expresión de los genes de globina. A. Los represores dCas9 y dCas9-KRAB se coexpresaron en un vector lentivírico con ARNg individuales. Cuando se analiza mediante RT-PCR cuantitativa 7 días después de la transducción, Los K562 que expresan dCas9-KRAB logran hasta un 95 % de represión de los genes de B. γ -globina, C. ϵ -globina y D. β -globina, en comparación con las células control que no recibieron tratamiento lentivírico.

30 35 La Fig. 58 muestra que el aislamiento del "núcleo" de HAT p300 para la modificación epigenética dirigida de histonas solo a través de la fusión dCas9.

40 La Fig. 59 muestra un esquema simplificado de *S. pyogenes* fusión dCas9-VP64 (arriba) y fusión del núcleo dCas9-p300 (abajo). Los motivos adyacentes del protoespaciador (PAM) se muestran con flechas en los loci del gen diana y el ARN guía sintético (ARNg) se muestra con flechas rayadas.

45 50 Las Fig. 60A-60C muestran datos representativos en tres loci humanos que demuestran la eficacia de la activación usando dCas9-p300 en relación con dCas9-VP64 y dCas9 sin ningún dominio efecto fusionado en la línea de cultivo celular 293T humana.

55 60 Las figuras 61A-61C muestran las secuencias de aminoácidos de las construcciones de dCas9. La leyenda de todas las Figs. 61A-61C se muestra en la Fig. 61A.

La Fig. 62 muestra que las proteínas de fusión HAT-dCase9-p300 no logran activar la expresión génica.

65 La Fig. 63 muestra que los ARNg también actúan sinérgicamente con el núcleo dCas9-p300.

La Fig. 64 muestra la integración mediada por TALEN de minidistrofina en la 5'UTR de la isoforma de distrofina del músculo esquelético Dp427m en líneas celulares de mioblastos esqueléticos derivadas de pacientes humanos con DMD que portan diferentes eliminaciones en el gen de la distrofina. Las células de pacientes con DMD se sometieron a electroporación con construcciones que codifican un par TALEN activo en el locus 5'UTR y una plantilla donante que porta el gen de minidistrofina, (a) Esquema que muestra cómo se integra la minidistrofina en la 5'UTR. (b) Se aislaron líneas celulares clonales resistentes a higromicina y se examinaron mediante PCR para detectar integraciones exitosas de sitios específicos en la 5'UTR usando los cebadores que se muestran en (a). Los asteriscos indican clones seleccionados para análisis adicionales en (c). (c) Se diferenciaron mioblastos DMD clonalmente aislados con eventos de integración detectados durante 6 días y se evaluó la expresión de una etiqueta HA fusionada al extremo C terminal de minidistrofina.

Descripción detallada

Como se describe en el presente documento, se ha descubierto que determinados métodos y composiciones de sistemas basados en CRISPR/Cas 9 asociado a CRISPR son útiles para alterar la expresión de genes, ingeniería genómica y corrección o reducción de los efectos de mutaciones en genes implicados en enfermedades genéticas. El sistema basado en CRISPR/Cas9 implica una proteína Cas9 y al menos un ARN guía, que proporciona la especificidad de dirección del ADN para el sistema. En particular, la presente divulgación describe una proteína de fusión Cas9 que combina la función de direccionamiento de secuencia de ADN del sistema basado en CRISPR/Cas9 con una actividad adicional, permitiendo de esta manera cambios en la expresión génica y/o el estado epigenético. El sistema también puede usarse en ingeniería genómica y en la corrección o reducción de los efectos de las mutaciones génicas.

La presente divulgación también proporciona determinadas composiciones y métodos para suministrar un sistema basado en CRISPR/asociado a CRISPR (Cas) 9 y múltiples ARNg para dirigirse a uno o más genes endógenos. La cotransfección de múltiples ARNm dirigidos a un único promotor permite la activación sinérgica, sin embargo, la

cotransfección de múltiples plásmidos conduce a niveles de expresión variables en cada célula debido a diferencias en el número de copias. Adicionalmente, la activación génica después de la transfección es transitoria debido a la dilución del ADN plasmídico con el tiempo. Por otra parte, muchos tipos de células no se transfecan fácilmente y la expresión génica transitoria puede no ser suficiente para inducir un efecto terapéutico. Para abordar estas limitaciones, 5 se desarrolló un sistema lentivírico individual para expresar Cas9 y hasta cuatro ARNgm de promotores independientes. Se divulga una plataforma que expresa proteínas de fusión Cas9 o dCas9 y hasta cuatro ARNg de un vector lentivírico individual. El vector lentivírico expresa un Cas9 o dCas9-VP64 constitutivo o inducible además de uno, dos, tres o cuatro ARNg expresados a partir de promotores independientes. Este sistema permite controlar tanto 10 la magnitud como el momento de la regulación génica basada en CRISPR/Cas9. Adicionalmente, la plataforma lentivírica proporciona niveles potentes y sostenidos de expresión génica que facilitarán las aplicaciones terapéuticas del sistema CRISPR/Cas9 en células primarias. Finalmente, este sistema puede usarse para editar múltiples genes simultáneamente, tal como la inactivación concomitante de varios oncogenes.

15 La presente divulgación también proporciona determinadas composiciones y métodos para suministrar nucleasas específicas de sitio al músculo esquelético y al músculo cardíaco usando vectores de virus adenoasociados (Adeno-Associated Virus, AAV) modificados. Las nucleasas específicas de sitio, que pueden modificarse por ingeniería genética, son útiles para alterar la expresión de genes, ingeniería genómica, corregir o reducir los efectos de mutaciones en genes implicados en enfermedades genéticas, o manipular genes implicados en otras afecciones que afectan el músculo esquelético o el músculo cardíaco o la regeneración muscular. Las nucleasas específicas de sitio 20 diseñadas pueden incluir una nucleasa con dedos de cinc (Zinc Finger Nuclease, ZFN), una nucleasa efectora TAL (TAL Effector Nuclease, TALEN) y/o un sistema CRISPR/Cas9 para la edición genómica. Como se describe en el presente documento, se editaron con éxito genes en el tejido del músculo esquelético *in vivo* usando este sistema de suministro único. En el presente documento se divulga un medio para reescribir el genoma humano para aplicaciones terapéuticas y especies modelo diana para aplicaciones científicas básicas.

25 La edición génica depende en gran medida del ciclo celular y de vías complejas de reparación del ADN que varían de un tejido a otro. El músculo esquelético es un entorno muy complejo, que consiste en grandes miofibras con más de 100 núcleos por célula. La terapia génica y la biología en general han estado limitadas durante décadas por obstáculos de suministro *in vivo*. Estos desafíos incluyen la estabilidad del vehículo *in vivo*, dirigirse al tejido correcto, obtener 30 suficiente expresión génica y producto genético activo y evitar la toxicidad que podría superar la actividad, lo cual es común con las herramientas de edición génica. Otros vehículos de suministro, tales como la inyección directa de ADN plasmídico, funcionan expresando genes en el músculo esquelético y el músculo cardíaco en otros contextos, pero no funcionan bien con estas nucleasas específicas de sitio para lograr niveles detectables de edición genómica.

35 Aunque muchas secuencias de genes son inestables en los vectores AAV y, por lo tanto, no pueden suministrarse, estas nucleasas específicas de sitio son sorprendentemente estables en los vectores AAV. Cuando estas nucleasas específicas de sitio se liberan y se expresan, permanecen activas en el tejido del músculo esquelético. La estabilidad de las proteínas y la actividad de las nucleasas específicas de sitio dependen en gran medida del tipo de tejido y del tipo de célula. Estas nucleasas activas y estables son capaces de modificar secuencias genéticas en el complejo 40 entorno del músculo esquelético. La divulgación actual describe una forma de suministrar formas activas de esta clase de terapia al músculo esquelético o al músculo cardíaco que es eficaz, eficiente y facilita la modificación exitosa del genoma.

45 La presente divulgación también proporciona determinadas moléculas efectoras epigenéticas de fusión, una proteína de fusión dCas9-p300, que proporciona una herramienta sólida y potencialmente más aplicable para la modulación transcripcional sintética en comparación con la fusión dCas9-VP64. Los genes diana activados en un grado sustancialmente mayor que la proteína de fusión dCas9-VP64 en todos los loci analizados, además, el p300 tiene actividad endógena intrínseca en potenciadores dentro del genoma humano. La proteína de fusión dCas9-p300 puede activar regiones potenciadoras y promotores de genes diana endógenos.

50 La proteína de fusión dCas9-p300 puede usarse en líneas celulares de cultivo de tejidos humanos para activar la expresión génica. Esta proteína de fusión puede usarse para dirigir el estado epigenético de loci diana dentro de las células humanas con precisión y previsibilidad para controlar la diferenciación, modular la regulación celular y aplicar 55 terapias potenciales innovadoras. Las tecnologías actuales están limitadas en cuanto a la fuerza de activación y el alcance y la sostenibilidad de la modulación epigenética; obstáculos que pueden evitarse mediante la utilización de esta nueva proteína de fusión.

Los encabezados de sección como se usan en esta sección y la divulgación completa tienen un fin sencillamente organizativo y no pretenden ser limitantes.

60

1. Definiciones

A menos que se definan de otra manera, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que el comúnmente entendido por un experto habitual en la materia. En caso de conflicto, el presente documento, incluyendo definiciones, controlará. Los métodos y materiales preferidos se describen más adelante, aunque en la práctica o ensayo de la presente invención pueden usarse métodos y materiales similares o

equivalentes a los descritos en la presente divulgación.

Las expresiones "comprende (comprenden)", "incluye (incluyen)", "que tiene", "tiene", "puede", "contiene (contienen)", y variantes de las mismas, como se usa en el presente documento, pretenden ser expresiones, términos o palabras

5 de transición abiertos que no excluyen la posibilidad de actos o estructuras adicionales. Las formas en singular "un/una", "y" y "el/la" incluyen las referencias en plural salvo que el contexto indique claramente otra cosa. La presente divulgación también contempla otras realizaciones "que comprenden", "que consisten en" y "que consisten esencialmente en", las realizaciones o elementos presentados en el presente documento, ya se expongan o no explícitamente.

10 Para la cita de intervalos numéricos en el presente documento, se contempla explícitamente cada número intermedio con el mismo grado de precisión. Por ejemplo, para el intervalo de 6-9, se contemplan los números 7 y 8 además de 6 y 9 y para el intervalo 6,0-7,0, se contemplan explícitamente los números 6,0, 6,1, 6,2, 6,3, 6,4, 6,5, 6,6, 6,7, 6,8, 6,9 y 7,0.

15 "Virus adenoasociado" o "AAV", como se usan indistintamente en el presente documento, se refiere a un pequeño virus que pertenece al género *Dependovirus* de la familia *Parvoviridae* que infecta a seres humanos y algunas otras especies de primates. Actualmente no se sabe que el AAV provoque enfermedades y, en consecuencia, el virus provoca una respuesta inmunitaria muy leve.

20 "Región de unión" como se usa en el presente documento se refiere a la región dentro de una región diana de nucleasa que es reconocida y unida por la nucleasa.

25 "Músculo cardíaco" o "músculo del corazón" como se usan indistintamente en el presente documento significa un tipo de músculo estriado involuntario que se encuentra en las paredes y la base histológica del corazón, el miocardio. El músculo cardíaco está formado por cardiomiocitos o miocardiocitos. Los miocardiocitos muestran estrías similares a las de las células del músculo esquelético, pero contienen solo un núcleo único, a diferencia de las células esqueléticas multinucleadas.

30 "Afección del músculo cardíaco" como se usa en el presente documento se refiere a una afección relacionada con el músculo cardíaco, tal como miocardiopatía, insuficiencia cardiaca, arritmia y enfermedades cardíacas inflamatorias.

35 "Secuencia codificante" o "ácido nucleico codificante" como se usa en el presente documento significa los ácidos nucleicos (molécula de ARN o ADN) que comprenden una secuencia nucleotídica que codifica una proteína. La secuencia codificante puede incluir además señales de iniciación y terminación unidas operativamente a elementos reguladores que incluyen un promotor y una señal de poliadenilación capaz de dirigir la expresión en las células de un individuo o mamífero al que se administra el ácido nucleico. La secuencia codificante puede ser con codones optimizados.

40 "Complemento" o "complementario" como se usa en el presente documento significa que un ácido nucleico puede significar emparejamiento de bases de Watson-Crick (por ejemplo, A-T/U y C-G) o de Hoogsteen entre nucleótidos o análogos de nucleótidos de moléculas de ácido nucleico. "Complementariedad" se refiere a una propiedad compartida entre dos secuencias de ácidos nucleicos, de tal manera que cuando están alineados antiparalelos entre sí, las bases de nucleótidos en cada posición serán complementarias.

45 "Corrección", "edición genómica" y "restauración" como se usan en el presente documento se refieren a cambiar un gen mutante que codifica una proteína truncada o ninguna proteína en absoluto, de tal manera que se obtenga una expresión de proteína funcional de longitud completa o parcialmente funcional de longitud completa. Corregir o restaurar un gen mutante puede incluir reemplazar la región del gen que tiene la mutación o reemplazar el gen mutante completo con una copia del gen que no tiene la mutación con un mecanismo de reparación tal como la reparación dirigida por homología (Homology-Directed Repair, HDR). Corregir o restaurar un gen mutante también puede incluir reparar una mutación de desplazamiento de marco que provoca un codón de parada prematuro, un sitio acceptor de corte y empalme aberrante o un sitio donante de corte y empalme aberrante, generando una rotura bicondicional en el gen que luego se repara mediante unión de extremos no homólogos (Non-Homologous End Joining, NHEJ). NHEJ puede añadir o eliminar al menos un par de bases durante la reparación lo que puede restaurar el marco de lectura adecuado y eliminar el codón de parada prematuro. Corregir o restaurar un gen mutante también puede incluir alterar un sitio acceptor de corte y empalme o una secuencia donante de corte y empalme aberrantes. Corregir o restaurar un gen mutante también puede incluir eliminar un segmento génico no esencial mediante la acción simultánea de dos nucleasas en la misma cadena de ADN para restaurar el marco de lectura adecuado eliminando el ADN entre los dos sitios diana de la nucleasa y reparando la ruptura de ADN mediante NHEJ.

60 "ADN donante", "plantilla donante" y "plantilla de reparación" como se usan indistintamente en el presente documento se refieren a un fragmento o molécula de ADN bicondicional que incluye al menos una porción del gen de interés. El ADN donante puede codificar una proteína completamente funcional o una proteína parcialmente funcional.

65 "Distrofia muscular de Duchenne" o "DMD" como se usan indistintamente en este documento se refieren a un trastorno

recesivo, mortal, ligado al cromosoma X que da como resultado degeneración muscular y final muerte. La DMD es una enfermedad monogénica hereditaria común y se produce en 1 de cada 3500 hombres. La DMD es el resultado de mutaciones heredadas o espontáneas que provocan mutaciones sin sentido o de cambio de marco en el gen de la distrofina. La mayoría de las mutaciones de la distrofina que provocan DMD son eliminaciones de exones que alteran el marco de lectura y provocan la terminación prematura de la traducción en el gen de la distrofina. Los pacientes con DMD normalmente pierden la capacidad de mantenerse físicamente a sí mismos durante la niñez, debilitándose progresivamente durante la adolescencia y mueren a los veinte años.

5 "Distrofina" como se usa en el presente documento se refiere a una proteína citoplasmática en forma de varilla que forma parte de un complejo proteico que conecta el citoesqueleto de una fibra muscular con la matriz extracelular circundante a través de la membrana celular. La distrofina proporciona estabilidad estructural al complejo de distroglicano de la membrana celular que es responsable de regular la integridad y la función de las células musculares. El gen de la distrofina o "gen DMD" como se usa indistintamente en el presente documento tiene 2,2 megabases en el locus Xp21. La transcripción primaria mide aproximadamente 2.400 kb y el siendo el ARNm maduro aproximadamente 14 kb. 79 exones codifican la proteína que tiene más de 3500 aminoácidos.

10 15 "Exón 51" como se usa en el presente documento se refiere al 51º exón del gen de la distrofina. El exón 51 suele estar adyacente a eliminaciones que alteran el marco en pacientes con DMD y ha sido la diana de ensayos clínicos para la omisión de exones basada en oligonucleótidos. Un ensayo clínico para el compuesto que omite el exón 51, eteplirsen, informó recientemente un beneficio funcional significativo durante 48 semanas, con un promedio del 47 % de fibras positivas para distrofina en comparación con el valor inicial. Las mutaciones en el exón 51 son ideales para la corrección permanente mediante la edición genómica basada en NHEJ.

20 25 "Desplazamiento de marco" o "mutación de desplazamiento de marco" como se usan indistintamente en el presente documento se refiere a un tipo de mutación genética en donde la adición o la eliminación de uno o más nucleótidos provoca un cambio en el marco de lectura de los codones en el ARNm. El desplazamiento en el marco de lectura puede conducir a la alteración en la secuencia de aminoácidos en la traducción de proteínas, tal como una mutación sin sentido o un codón de parada prematuro.

30 "Funcional" y "completamente funcional" como se usan en el presente documento describen una proteína que tiene actividad biológica. Un "gen funcional" se refiere a un gen transcripto a ARNm, que se traduce en una proteína funcional.

35 "Proteína de fusión" como se usa en el presente documento se refiere a una proteína químérica creada a través de la unión de dos o más genes que originalmente codificaban proteínas separadas. La traducción del gen de fusión da como resultado un único polipéptido con propiedades funcionales derivadas de cada una de las proteínas originales.

40 "Construcción genética" como se usa en el presente documento se refiere a las moléculas de ADN o ARN que comprenden una secuencia nucleotídica que codifica una proteína. La secuencia codificante incluye señales de iniciación y terminación unidas operativamente a elementos reguladores que incluyen un promotor y una señal de poliadensilación capaz de dirigir la expresión en las células del individuo a quien se administra la molécula de ácido nucleico. Como se usa en el presente documento, la expresión "forma expresable" se refiere a construcciones génicas que contienen los elementos reguladores necesarios operables unidos a una secuencia codificante que codifica una proteína de tal manera que cuando está presente en la célula del individuo, se expresará la secuencia codificante.

45 50 "Enfermedad genética" como se usa en el presente documento se refiere a una enfermedad, parcial o completamente, directa o indirectamente, provocada por una o más anomalías en el genoma, especialmente una afección que está presente desde el nacimiento. La anomalía puede ser una mutación, una inserción o una eliminación. La anomalía puede afectar a la secuencia codificante del gen o su secuencia reguladora. La enfermedad genética puede ser, pero no limitada a DMD, hemofilia, fibrosis quística, corea de Huntington, hipercolesterolemia familiar (defecto del receptor de LDL), hepatoblastoma, enfermedad de Wilson, porfiria hepática congénita, trastornos hereditarios del metabolismo hepático, Síndrome de Lesch Nyhan, anemia de células falciformes, talasemias, xeroderma pigmentoso, anemia de Fanconi, retinitis pigmentosa, ataxia telangiectasia, síndrome de Bloom, retinoblastoma y enfermedad de Tay-Sachs.

55 60 "Reparación dirigida por homología" o "HDR" como se usan indistintamente en el presente documento se refiere a un mecanismo en las células para reparar lesiones de ADN bicatenario cuando una parte homóloga de ADN está presente en el núcleo, principalmente en las fases G2 y S del ciclo celular. HDR usa una plantilla de ADN de un donante para guiar la reparación y puede usarse para crear cambios de secuencia específicos en el genoma, incluyendo la adición dirigida de genes completos. Si se proporciona una plantilla donante junto con la nucleasa específica del sitio, tal como con sistemas basados en CRISPR/Cas9, entonces la maquinaria celular reparará la rotura mediante recombinación homóloga, que se mejora varios órdenes de magnitud en presencia de escisión del ADN. Cuando la parte de ADN homólogo está ausente, en su lugar puede realizarse una unión de extremos no homólogos.

65 "Edición genómica" como se usa en el presente documento se refiere a cambiar un gen. La edición genómica puede incluir la corrección o la restauración de un gen mutante. La edición genómica puede incluir la inactivación de un gen, tal como un gen mutante o un gen normal. La edición genómica puede usarse para tratar enfermedades o mejorar la reparación muscular cambiando el gen de interés.

- "Idéntico" o "identidad" como se usa en el presente documento en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas significa que las secuencias tienen un porcentaje específico de residuos que son iguales en una región específica. El porcentaje puede calcularse alineando óptimamente las dos secuencias, comparando las 5 dos secuencias a lo largo de la región especificada, determinando el número de posiciones en las que aparece el resto idéntico en ambas secuencias para proporcionar el número de posiciones emparejadas, dividiendo el número de posiciones emparejadas entre el número total de posiciones en la región especificada y multiplicando el resultado por 100 para dar el porcentaje de identidad de secuencia. En los casos donde las dos secuencias tienen longitudes 10 diferentes o el alineamiento produce uno o más extremos escalonados y la región de comparación especificada incluye únicamente una sola secuencia, los restos de una secuencia individual se incluyen en el denominador pero no en el numerador del cálculo. Al comparar ADN y ARN, la timina (T) y el uracilo (U) pueden considerarse equivalentes. La identidad puede realizarse manualmente o mediante el uso de un algoritmo de secuencia informática tales como BLAST o BLAST 2.0.
- 15 "Gen mutante" o "gen mutado" como se usan indistintamente en el presente documento se refiere a un gen que se ha sometido a una mutación detectable. Un gen mutante ha sufrido un cambio, tal como la pérdida, la ganancia o el intercambio de material genético, lo que afecta la transmisión y expresión normal del gen. Un "gen alterado" como se usa en el presente documento se refiere a un gen mutante que tiene una mutación que provoca un codón de parada prematuro. El producto génico alterado se trunca con respecto a un producto génico no alterado de longitud completa.
- 20 La "vía de unión de extremos no homólogos (NHEJ)" como se usa en el presente documento se refiere a una vía que repara roturas biconatenarias en el ADN ligando directamente los extremos rotos sin la necesidad de una plantilla homóloga. La religación independiente de la plantilla de los extremos del ADN mediante NHEJ es un proceso de reparación estocástico, propenso a errores que introduce microinsertiones y microeliminaciones aleatorias (indeles) 25 en el punto de ruptura del ADN. Este método puede usarse para, intencionalmente, interrumpir, eliminar o alterar el marco de lectura de secuencias genéticas específicas. NHEJ normalmente usa secuencias cortas de ADN homólogas llamadas microhomologías para guiar la reparación. Estas microhomologías suelen estar presentes en salientes monocatenarios al final de las roturas biconatenarias. Cuando los salientes son perfectamente compatibles, NHEJ suele reparar la rotura con precisión, sin embargo, también puede producirse una reparación imprecisa que conduzca a la pérdida de nucleótidos, pero es mucho más común cuando los salientes no son compatibles.
- 30 "Gen normal" como se usa en el presente documento se refiere a un gen que no se ha sometido a un cambio, tal como una pérdida, una ganancia o un intercambio de material genético. El gen normal se somete a una transmisión y expresión génica normales.
- 35 "NHEJ mediado por nucleasa" como se usa en el presente documento se refiere a NHEJ que se inicia después de que una nucleasa, tal como una cas9, corte el ADN biconatenario.
- 40 "Ácido nucleico", "oligonucleótido" o "polinucleótido" como se usa en el presente documento significa al menos dos nucleótidos unidos covalentemente entre sí. La representación de una única cadena también define la secuencia de la cadena complementaria. Por lo tanto, un ácido nucleico también abarca la cadena complementaria de una cadena única representada. Pueden usarse muchas variantes de un ácido nucleico para el mismo fin que un ácido nucleico determinado. Por lo tanto, un ácido nucleico también abarca ácidos nucleicos sustancialmente idénticos y 45 complementos de los mismos. Una cadena sencilla proporciona una sonda que puede hibridarse con una secuencia diana en condiciones de hibridación estrictas. Por lo tanto, un ácido nucleico también abarca una sonda que se hibrida en condiciones de hibridación estrictas.
- 50 Los ácidos nucleicos pueden ser monocatenarios o biconatenarios, o pueden contener porciones de secuencias tanto biconatenarias como monocatenarias. El ácido nucleico puede ser ADN, tanto genómico como de ADNc, ARN o un híbrido, donde el ácido nucleico puede contener combinaciones de desoxirribo y ribonucleótidos, y combinaciones de bases que incluyen uracilo, adenina, timina, citosina, guanina, inosina, xantina, hipoxantina, isocitosina e isoguanina. Los ácidos nucleicos pueden obtenerse mediante métodos de síntesis química o mediante métodos recombinantes.
- 55 "Unido operativamente" como se usa en el presente documento significa que la expresión de un gen está bajo el control de un promotor con el que está conectado espacialmente. Un promotor puede estar colocado en 5' (en dirección 5') o 3' (en dirección 3') de un gen bajo su control. La distancia entre el promotor y un gen puede ser aproximadamente la misma que la distancia entre ese promotor y el gen que controla en el gen del que deriva el promotor. Como se conoce en la técnica, la variación en esta distancia puede adaptarse sin pérdida de la función promotora.
- 60 "Parcialmente funcional" como se usa en el presente documento describe una proteína que está codificada por un gen mutante y tiene menos actividad biológica que una proteína funcional pero más que una proteína no funcional.
- 65 "Codón de parada prematuro" o "codón de parada fuera de marco" como se usan indistintamente en el presente documento se refiere a una mutación sin sentido en una secuencia de ADN, que da como resultado un codón de parada en una ubicación que normalmente no se encuentra en el gen de tipo silvestre. Un codón de parada prematuro puede hacer que una proteína se trunque o sea más corta en comparación con la versión de longitud completa de la

proteína.

"Promotor" como se usa en el presente documento significa una molécula sintética o de origen natural que es capaz de conferir, activar o mejorar la expresión de un ácido nucleico en una célula. Un promotor puede comprender una o

- 5 más secuencias reguladoras transcripcionales específicas para mejorar aún más la expresión y/o alterar la expresión espacial y/o la expresión temporal de las mismas. Un promotor también puede comprender elementos potenciadores o represores distales, que pueden localizarse hasta varios miles de pares de bases desde el sitio de inicio de la transcripción. Un promotor puede derivar de fuentes que incluyen víricas, bacterianas, fúngicas, de plantas, de insectos y de animales. Un promotor puede regular la expresión de un componente genético de manera constitutiva o diferencial
- 10 con respecto a la célula, el tejido o el órgano donde se produce la expresión o, con respecto a la etapa de desarrollo en la que se produce la expresión, o en respuesta a estímulos externos tales como estreses fisiológicos, patógenos, iones metálicos o agentes inductores. Los ejemplos representativos de promotores incluyen el promotor del bacteriófago T7, el promotor del bacteriófago T3, el promotor SP6, el operador-promotor lac, el promotor tac, el promotor tardío de SV40, el promotor temprano de SV40, el promotor de RSV-LTR, el promotor de CMV IE, el promotor temprano de SV40 o el promotor tardío de SV40 y el promotor de CMV IE.
- 15

"Dirresiduo variable repetido" o "RVD" como se usan indistintamente en el presente documento se refiere a un par de restos de aminoácidos adyacentes dentro de un motivo de reconocimiento de ADN (también conocido como "módulo RVD"), que incluye 33-35 aminoácidos, de un dominio de unión al ADN de TALE. El RVD determina la especificidad

- 20 de nucleótidos del módulo RVD. Los módulos RVD pueden combinarse para producir una matriz RVD. La "longitud de la matriz RVD" como se usa en el presente documento se refiere al número de módulos RVD que corresponde a la longitud de la secuencia nucleotídica dentro de la región diana de TALEN que es reconocida por un TALEN, es decir, la región de unión.

- 25 "Nucleasa específica de sitio" como se usa en el presente documento se refiere a una enzima capaz de reconocer y escindir específicamente secuencias de ADN. La nucleasa específica de sitio puede diseñarse por ingeniería genética. Los ejemplos de nucleasas específicas de sitio diseñadas incluyen nucleasas con dedos de cinc (ZFN), nucleasas efectoras TAL (TALEN) y sistemas basados en CRISPR/Cas9.

- 30 "Músculo esquelético" como se usa en el presente documento se refiere a un tipo de músculo estriado, que está bajo el control del sistema nervioso somático y unido a los huesos mediante haces de fibras de colágeno conocidos como tendones. El músculo esquelético está formado por componentes individuales conocidos como miocitos o "células musculares", a veces se les llama coloquialmente "fibras musculares". Los miocitos se forman a partir de la fusión de mioblastos del desarrollo (un tipo de célula progenitora embrionaria que da lugar a una célula muscular) en un proceso conocido como miogénesis. Estas células largas, cilíndricas, multinucleadas también se llaman miofibras.
- 35

"Afección del músculo esquelético" como se usa en el presente documento se refiere a una afección relacionada con el músculo esquelético, tales como distrofias musculares, envejecimiento, degeneración muscular, cicatrización de heridas y debilidad o atrofia muscular.

- 40 "Espaciadores" y "región espaciadora" como se usan indistintamente en el presente documento se refieren a la región dentro de una región diana TALEN o ZFN que está entre, pero no es parte de, las regiones de unión para dos TALEN o ZFN.

- 45 "Sujeto" y "paciente" como se usan en el presente documento se refieren indistintamente a cualquier vertebrado, incluyendo, pero sin limitación, un mamífero (por ejemplo, vaca, cerdo, camello, llama, caballo, cabra, conejo, oveja, hámsteres, cobaya, gato, perro, rata y ratón, un primate no humano (por ejemplo, un mono, tal como un macaco cangrejero o un macaco de la India, chimpancé, etc.) y un ser humano). En algunas realizaciones, el sujeto puede ser un ser humano o un animal no humano. El sujeto o paciente puede estar sometiéndose a otras formas de tratamiento.
- 50

"Gen diana" como se usa en el presente documento se refiere a cualquier secuencia nucleotídica que codifique un producto génico conocido o putativo. El gen diana puede ser un gen mutado implicado en una enfermedad genética.

- 55 "Región diana" como se usa en el presente documento se refiere a la región del gen diana a la que la nucleasa específica del sitio está diseñada para unirse y escindir.

- 60 "Efecto similar a un activador de la transcripción" o "TALE" como se usa en el presente documento se refiere a una estructura proteica que reconoce y se une a una secuencia de ADN particular. El "dominio de unión a ADN de TALE" se refiere a un dominio de unión a ADN que incluye una serie de repeticiones de aminoácidos en tandem de 33-35, también conocidos como módulos RVD, cada uno de los cuales reconoce específicamente un único par de bases de ADN. Los módulos RVD pueden disponerse en cualquier orden para ensamblar una matriz que reconozca una secuencia definida.

- 65 Una especificidad de unión de un dominio de unión a ADN de TALE se determina mediante la matriz RVD seguida de una única repetición truncada de 20 aminoácidos. Un dominio de unión al ADN de TALE puede tener de 12 a 27 módulos RVD, cada uno de los cuales contiene un RVD y reconoce un único par de bases de ADN. Se han

- identificado RVD específicos que reconocen cada uno de los cuatro posibles nucleótidos del ADN (A, T, C y G). Debido a que los dominios de unión al ADN de TALE son modulares, las repeticiones que reconocen los cuatro nucleótidos de ADN diferentes pueden unirse entre sí para reconocer cualquier secuencia de ADN en particular. Estos dominios de unión al ADN específicos pueden combinarse después con dominios catalíticos para crear enzimas funcionales, incluyendo factores de transcripción artificiales, metiltransferasas, integrasas, nucleasas y recombinasas.
- "Nucleasas efectoras similares a activadores de la transcripción" o "TALEN" como se usan indistintamente en el presente documento se refieren a proteínas de fusión diseñadas del dominio catalítico de una nucleasa, tal como la endonucleasa *FokI* y un dominio de unión al ADN TALE diseñado que puede dirigirse a una secuencia de ADN personalizada. Un "monómero TALEN" se refiere a una proteína de fusión diseñada con un dominio de nucleasa catalítica y un dominio de unión al ADN TALE diseñado. Pueden diseñarse dos monómeros de TALEN para dirigirse a y escindir una región diana de TALEN.
- "Transgén" como se usa en el presente documento se refiere a un gen o material genético que contiene una secuencia génica que se ha aislado de un organismo y se introduce en un organismo diferente. Este segmento no nativo de ADN puede conservar la capacidad de producir ARN o proteínas en el organismo transgénico o puede alterar la función normal del código genético del organismo transgénico. La introducción de un transgén tiene el potencial de cambiar el fenotipo de un organismo.
- "Variante" usada en el presente documento con respecto a un ácido nucleico significa (i) una porción o fragmento de una secuencia nucleotídica de referencia; (ii) el complemento de una secuencia nucleotídica de referencia o porción de la misma; (iii) un ácido nucleico que es sustancialmente idéntico a un ácido nucleico de referencia o su complemento; o (iv) un ácido nucleico que se hibrida en condiciones estrictas con el ácido nucleico de referencia, el complemento del mismo o una secuencia sustancialmente idéntica al mismo.
- "Variante" con respecto a un péptido o polipéptido que difiere en la secuencia de aminoácidos por la inserción, eliminación o sustitución conservativa de aminoácidos, pero conserva al menos una actividad biológica. Variante también puede significar una proteína con una secuencia de aminoácidos que es sustancialmente idéntica a una proteína de referencia con una secuencia de aminoácidos que conserva al menos una actividad biológica. Una sustitución conservativa de un aminoácido, es decir, el reemplazo de un aminoácido por un aminoácido diferente con propiedades similares (por ejemplo, hidrofilia, grado y distribución de regiones cargadas) se reconoce en la técnica implicando normalmente un cambio menor. Estos cambios menores pueden identificarse, en parte, tomando en consideración el índice hidropático de los aminoácidos, como se entiende en la técnica Kyte *et al.*, *J. Mol. Biol.* 157:105-132 (1982). El índice hidropático de un aminoácido se basa en una consideración de su hidrofobia y carga. Se sabe en la técnica que pueden sustituirse aminoácidos de índices hidropáticos similares y aún conservar la función proteica. En un aspecto, se sustituyen los aminoácidos que tienen índices hidropáticos de ± 2 . La hidrofilia de los aminoácidos también puede usarse para revelar sustituciones que darían como resultado que las proteínas conserven su función biológica. Una consideración de la hidrofilia de los aminoácidos en el contexto de un péptido permite calcular la mayor hidrofilia promedio local de ese péptido. Pueden llevarse a cabo sustituciones con aminoácidos que tienen valores de hidrofilia de ± 2 entre sí. Tanto el índice de hidrofilia como el valor de hidrofilia de los aminoácidos están influenciados por la cadena lateral particular de dicho aminoácido. De manera coherente con esta observación, se entiende que las sustituciones de aminoácidos que son comparables con la función biológica dependen de la similitud relativa de los aminoácidos y en particular, las cadenas laterales de estos aminoácidos, como se revela por la hidrofobia, hidrofilia, carga, tamaño y otras propiedades.
- "Vector" como se usa en el presente documento significa una secuencia de ácido nucleico que contiene un origen de replicación. Un vector puede ser un vector vírico, un bacteriófago, un cromosoma artificial bacteriano o un cromosoma artificial de levadura. Un vector puede ser un vector de ADN o ARN. Un vector puede ser un vector extracromosómico autorreplicante y, preferentemente, es un plásmido de ADN. Por ejemplo, el vector puede codificar una proteína de fusión iCas9-VP64 que comprende la secuencia de aminoácidos de SEQ ID NO: 1 o al menos una secuencia nucleotídica de ARNg de una cualquiera de las SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625. Como alternativa, el vector puede codificar Cas9 y al menos una secuencia nucleotídica de ARNg de cualquiera de las SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625.
- "Dedo de cinc" como se usa en el presente documento se refiere a una estructura proteica que reconoce y se une a secuencias de ADN. El dominio de dedos de cinc es el motivo de unión al ADN más común en el proteoma humano. Un solo dedo de cinc contiene aproximadamente 30 aminoácidos y el dominio normalmente funciona uniendo 3 pares de bases consecutivos de ADN mediante interacciones de una única cadena lateral de aminoácido por par de bases.
- "Nucleasa con dedos de cinc" o "ZFN" como se usan indistintamente en el presente documento se refiere a una molécula de proteína químérica que comprende al menos un dominio de unión al ADN con dedos de cinc efectivamente unido a al menos una nucleasa o parte de una nucleasa capaz de escindir el ADN cuando está completamente ensamblado.
- A menos que se defina lo contrario en el presente documento, los términos científicos y técnicos usados en relación con la presente divulgación han de tener los significados que se entienden comúnmente por los expertos habituales

en la materia. Por ejemplo, cualquier nomenclatura usada junto con, y las técnicas de, cultivo celular y cultivo de tejidos, biología molecular, inmunología, microbiología, genética y química e hibridación de proteínas y de ácidos nucleicos descritas en el presente documento son aquellas que son bien conocidas y comúnmente usadas en la técnica. El significado y el alcance de los términos ha de ser claro; en el caso, sin embargo, de cualquier ambigüedad

5 latente, las definiciones proporcionadas en el presente documento tienen precedencia sobre cualquier diccionario o definición extrínsecos. Además, a menos que el contexto requiera otra cosa, los términos en singular incluirán los plurales y los términos en plural incluirán el singular.

2. Composiciones para la edición genómica

10 La presente divulgación se dirige a composiciones para la edición genómica, la alteración genómica o la alteración de la expresión génica de un gen diana. Las composiciones pueden incluir un vector vírico y una proteína de fusión tal como una nucleasa específica de sitio o un sistema CRISPR/Cas9 con al menos un ARNg.

15 **a. Composiciones para la edición genómica en el músculo**

La presente divulgación se dirige a una composición para editar el genoma de un gen diana en el músculo esquelético o el músculo cardíaco de un sujeto. La composición incluye un vector AAV modificado y una secuencia nucleotídica que codifica una nucleasa específica de sitio. La composición suministra formas activas de nucleasas específicas de sitio al músculo esquelético o al músculo cardíaco. La composición puede comprender además un ADN donante o un transgén. Estas composiciones pueden usarse en edición genómica, ingeniería genómica y corrección o reducción de los efectos de mutaciones en genes implicados en enfermedades genéticas y/u otras afecciones del músculo esquelético o cardíaco.

25 El gen diana puede estar implicado en la diferenciación de una célula o en cualquier otro proceso donde la activación, la represión o la alteración de un gen puede desearse, o puede tener una mutación tal como una eliminación, una mutación de desplazamiento de marco o una mutación sin sentido. Si el gen diana tiene una mutación que provoca un codón de parada prematuro, un sitio aceptor de corte y empalme aberrante o un sitio donante de corte y empalme aberrante, las nucleasas específicas de sitio pueden diseñarse para reconocer y unirse a una secuencia nucleotídica 30 en dirección 5' o en dirección 3' del codón de parada prematuro, el sitio aceptor de corte y empalme aberrante o el sitio donante de corte y empalme aberrante. Las nucleasas específicas de sitio también pueden usarse para alterar el corte y empalme normal de genes dirigiéndose a los aceptores y donantes del corte y empalme para inducir la omisión de codones de parada prematuros o restaurar un marco de lectura alterado. Las nucleasas específicas de sitio pueden 35 mediar o no cambios inespecíficos en regiones del genoma que codifican proteínas.

3. Sistema CRISPR

"Grupos de repeticiones palindrómicas en intervalos regulares" y "CRISPR", como se usa indistintamente en el 40 presente documento, se refiere a loci que contienen múltiples repeticiones directas cortas que se encuentran en los genomas de aproximadamente el 40 % de las bacterias secuenciadas y el 90 % de las arqueas secuenciadas. El sistema CRISPR es un sistema de nucleasa microbiana implicado en la defensa contra fagos y plásmidos invasores que proporciona una forma de inmunidad adquirida. Los loci CRISPR en hospedadores microbianos contienen una combinación de genes asociados a CRISPR (Cas), así como elementos de ARN no codificantes capaces de programar la especificidad de la escisión de ácidos nucleicos mediada por CRISPR. Los segmentos cortos de ADN extraño, 45 llamados espaciadores, se incorporan al genoma entre repeticiones CRISPR y sirven como una "memoria" de exposiciones pasadas. Cas9 forma un complejo con el extremo 3' del ARNgm y el par proteína-ARN reconoce su diana genómica mediante el emparejamiento de bases complementarias entre el extremo 5' de la secuencia del ARNgm y una secuencia de ADN predefinida de 20 pb, conocida como el protoespaciador. Este complejo se dirige a loci homólogos del ADN del patógeno a través de regiones codificadas dentro del ARNcr, es decir, los protoespaciadores 50 y los motivos adyacentes a protoespaciadores (PAM) dentro del genoma del patógeno. La matriz CRISPR no codificante se transcribe y se escinde mediante repeticiones directas en ARNcr cortos que contienen secuencias espaciadoras individuales, que dirigen las nucleasas Cas al sitio diana (protoespaciador). Simplemente intercambiando la secuencia de reconocimiento de 20 pb del ARNgm expresado, la nucleasa Cas9 puede dirigirse a nuevas dianas genómicas. Los espaciadores CRISPR se usan para reconocer y silenciar elementos genéticos exógenos de manera 55 análoga a la iARN en organismos eucariotas.

Se conocen tres clases de sistemas CRISPR (sistemas efectores Tipos I, II y III). El sistema efector Tipo II lleva a cabo una rotura bicatenaria de ADN dirigida en cuatro etapas secuenciales, usando una única enzima efectora, Cas9, para escindir el ADNbc. En comparación con los sistemas efectores Tipo I y Tipo III, que requieren múltiples efectores 60 distintos que actúen como un complejo, el sistema efector de Tipo II puede funcionar en contextos alternativos tales como células eucariotas. El sistema efector Tipo II consiste en un largo pre-ARNcr, que se transcribe del locus CRISPR que contiene el espaciador, la proteína Cas9 y un ARNtracr, que participa en el procesamiento del pre-ARNcr. Los ARNtracr se hibridan con las regiones repetidas que separan los espaciadores del pre-ARNcr, iniciando de esta manera la escisión del ARNbc por la RNasa III endógena. Esta escisión es seguida por un segundo evento de escisión dentro de cada espaciador por Cas9, produciendo ARNcr maduros que permanecen asociados al ARNtracr y Cas9, formando un complejo Cas9:ARNcr-ARNtracr.

El complejo Cas9:ARNcr-ARNtracr desenrolla el dúplex de ADN y busca secuencias que coincidan con el ARNcr que se va a escindir. El reconocimiento de la diana se produce tras la detección de complementariedad entre una secuencia "protoespaciadora" en el ADN diana y la secuencia espaciadora restante en el ARNcr. Cas9 media la escisión del ADN diana si también está presente un motivo adyacente al protoespaciador (PAM) correcto en el extremo 3' del protoespaciador. Para dirigirse al protoespaciador, la secuencia debe ir seguida inmediatamente por el motivo adyacente al protoespaciador (PAM), una secuencia corta reconocida por la nucleasa Cas9 que se requiere para la escisión del ADN. Los diferentes sistemas de Tipo II tienen diferentes requisitos de PAM. El sistema CRISPR de *S. pyogenes* puede tener la secuencia PAM para este Cas9 (SpCas9) como 5'-NRG-3', donde R es cualquiera de A o G y caracterizar la especificidad de este sistema en células humanas. Una capacidad única del sistema CRISPR/Cas9 es la capacidad sencilla de dirigirse simultáneamente a múltiples loci genómicos distintos mediante la coexpresión de una única proteína Cas9 con dos o más ARNgs. Por ejemplo, el sistema Tipo II de *Streptococcus pyogenes* naturalmente prefiere usar una secuencia "NGG", donde "N" puede ser cualquier nucleótido, pero también acepta otras secuencias PAM, tal como "NAG" en sistemas modificados por ingeniería genética (Hsu *et al.*, *Nature Biotechnology* (2013) doi:10.1038/nbt.2647). De forma similar, el Cas9 derivado de *Neisseria meningitidis* (NmCas9) normalmente tiene un PAM nativo de NNNNGATT, pero tiene actividad a lo largo de una diversidad de PAM, incluyendo un PAM NNNNGNNN altamente degenerado (Esvelt *et al.* *Nature Methods* (2013) doi:10.1038/nmeth.2681).

4. Sistema basado en CRISPR/Cas9

Una forma modificada por ingeniería genética del sistema efector Tipo II de *Streptococcus pyogenes* se demostró que funciona en células humanas para la ingeniería genómica. En este sistema, la proteína Cas9 se dirigió a sitios genómicos diana mediante un "ARN guía" reconstituido sintéticamente ("ARNg", también se usa indistintamente en este documento como un ARN guía monocatenario químérico ("ARNgm")), que es una fusión ARNcr-ARNtracr que elimina la necesidad de procesamiento de RNasa III y ARNcr en general (véase la Fig. 53A). En el presente documento se proporcionan sistemas de ingeniería genética basados en CRISPR/Cas9 para su uso en la edición genómica y el tratamiento de enfermedades genéticas. Los sistemas modificados por ingeniería genética con CRISPR/Cas9 pueden diseñarse para dirigirse a cualquier gen, incluyendo genes implicados en una enfermedad genética, envejecimiento, regeneración de tejidos o cicatrización de heridas. Los sistemas basados en CRISPR/Cas9 pueden incluir una proteína Cas9 o una proteína de fusión Cas9 y al menos un ARNg. La proteína de fusión Cas9 puede, por ejemplo, incluir un dominio que tiene una actividad diferente a la endógena de Cas9, tal como un dominio de transactivación.

El gen diana puede estar implicado en la diferenciación de una célula o en cualquier otro proceso donde pueda desearse la activación de un gen o puede tener una mutación tal como una mutación por desplazamiento de marco o una mutación sin sentido. Si el gen diana tiene una mutación que provoca un codón de parada prematuro, un sitio acceptor de corte y empalme aberrante o un sitio donante de corte y empalme aberrante, el sistema basado en CRISPR/Cas9 puede diseñarse para reconocer y unirse a una secuencia nucleotídica en dirección 5' o en dirección 3' del codón de parada prematuro, el sitio acceptor de corte y empalme aberrante o el sitio donante de corte y empalme aberrante. El sistema basado en CRISPR-Cas9 también puede usarse para interrumpir el corte y empalme normal de genes dirigiéndose a los aceptores y donantes de corte y empalme para inducir la omisión de codones de parada prematuros o restaurar un marco de lectura alterado. El sistema basado en CRISPR/Cas9 puede mediar o no cambios inespecíficos en las regiones codificantes de proteínas del genoma.

a. Cas9

El sistema basado en CRISPR/Cas9 puede incluir una proteína Cas9 o una proteína de fusión Cas9. La proteína Cas9 es una endonucleasa que escinde el ácido nucleico y está codificada por los loci CRISPR y participa en el sistema CRISPR Tipo II. La proteína Cas9 puede ser de cualquier especie de bacteria o arquea, tal como *Streptococcus pyogenes*. La proteína Cas9 puede mutarse de tal manera que se inactive la actividad nucleasa. Una proteína Cas9 inactivada de *Streptococcus pyogenes* (iCas9, también denominada "dCas9") sin actividad endonucleasa se ha dirigido recientemente a genes en bacterias, levadura y células humanas mediante ARNg para silenciar la expresión génica a través del impedimento estérico. Como se usa en el presente documento, tanto "iCas9" como "dCas9" se refieren a una proteína Cas9 que tiene las sustituciones de aminoácidos D10A y H840A y tiene su actividad nucleasa inactivada. Por ejemplo, el sistema basado en CRISPR/Cas9 puede incluir una Cas9 de SEQ ID NO: 459 o 461.

b. Proteína de fusión Cas9

El sistema basado en CRISPR/Cas9 puede incluir una proteína de fusión. La proteína de fusión puede comprender dos dominios polipeptídicos heterólogos, en donde el primer dominio polipeptídico comprende una proteína Cas y el segundo dominio polipeptídico tiene una actividad tal como actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa o actividad desmetilasa. La proteína de fusión puede incluir una proteína Cas9 o una proteína Cas9 mutada, como se ha descrito anteriormente, fusionada a un segundo dominio polipeptídico que tiene una actividad tal como actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa o actividad

desmetilasa.

(1) Actividad de activación de la transcripción

- 5 El segundo dominio polipeptídico puede tener actividad de activación de la transcripción, es decir, un dominio de transactivación. Por ejemplo, la expresión génica de genes endógenos de mamíferos, tales como genes humanos, puede lograrse dirigiendo una proteína de fusión de iCas9 y un dominio de transactivación a promotores de mamíferos mediante combinaciones de ARNg. El dominio de transactivación puede incluir una proteína VP16, múltiples proteínas VP16, tales como un dominio VP48 o un dominio VP64 o un dominio p65 de la actividad activadora de la transcripción de NF kappa B. Por ejemplo, la proteína de fusión puede ser iCas9-VP64.
- 10

(2) Actividad de represión de la transcripción

- 15 El segundo dominio polipeptídico puede tener actividad de represión de la transcripción. El segundo dominio polipeptídico puede tener una actividad de caja asociada a Kruppel, tal como un dominio KRAB, actividad del dominio represor de ERF, actividad del dominio represor de MxI, actividad del dominio represor de SID4X, actividad del dominio represor de Mad-SID o actividad de la proteína de unión a la caja TATA. Por ejemplo, la proteína de fusión puede ser dCas9-KRAB.

- 20 **(3) Actividad del factor de liberación de transcripción**

El segundo dominio polipeptídico puede tener actividad de factor de liberación de la transcripción. El segundo dominio polipeptídico puede tener actividad del factor de liberación eucariótico 1 (ERF1) o actividad del factor de liberación eucariótico 3 (ERF3).

- 25 **(4) Actividad de modificación de histonas**

El segundo dominio polipeptídico puede tener actividad de modificación de histonas. El segundo dominio polipeptídico puede tener actividad histona desacetilasa, histona acetiltransferasa, histona desmetilasa o histona metiltransferasa.

- 30 La histona acetiltransferasa puede ser p300 o proteína de unión a CREB (CBP), o fragmentos de las mismas. Por ejemplo, la proteína de fusión puede ser dCas9-p300.

(5) Actividad nucleasa

- 35 El segundo dominio polipeptídico puede tener actividad nucleasa que es diferente de la actividad nucleasa de la proteína Cas9. Una nucleasa, o una proteína que tiene actividad nucleasa, es una enzima capaz de escindir los enlaces fosfodiéster entre las subunidades de nucleótidos de los ácidos nucleicos. Las nucleasas generalmente se dividen en endonucleasas y exonucleasas, aunque algunas de las enzimas pueden pertenecer a ambas categorías. Las nucleasas más conocidas son la desoxirribonucleasa y la ribonucleasa.
- 40

(6) Actividad de asociación de ácidos nucleicos

- 45 El segundo dominio polipeptídico puede tener actividad de asociación de ácido nucleico o proteína de unión a ácido nucleico. El dominio de unión a ADN (DNA-binding Domain, DBD) es un dominio proteico plegado independientemente que contiene al menos un motivo que reconoce ADN monocatenario o bicatenario. Un DBD puede reconocer una secuencia de ADN específica (una secuencia de reconocimiento) o tener una afinidad general por el ADN, región de asociación de ácido nucleico seleccionada del grupo que consiste en la región hélice-giro-hélice, región de cremallera de leucina, región de hélice alada, región de hélice-giro-hélice alada, región hélice-lazo-hélice, plegamiento de inmunoglobulina, dominio B3, Dedo de cinc, caja HMG, dominio Woz3, dominio de unión al ADN efector TAL.
- 50

(7) Actividad metilasa

- 55 El segundo dominio polipeptídico puede tener actividad metilasa, que implica transferir un grupo metilo al ADN, ARN, proteína, molécula pequeña, citosina o adenina. El segundo dominio polipeptídico puede incluir una ADN metiltransferasa.

(8) Actividad desmetilasa

- 60 El segundo dominio polipeptídico puede tener actividad desmetilasa. El segundo dominio polipeptídico puede incluir una enzima que elimina grupos metilo (CH₃-) de ácidos nucleicos, proteínas (en particular histonas) y otras moléculas. Como alternativa, el segundo polipéptido puede convertir el grupo metilo en hidroximetilcitosina en un mecanismo para desmetilar el ADN. El segundo polipéptido puede catalizar esta reacción. Por ejemplo, el segundo polipéptido que cataliza esta reacción puede ser Tet1.

- 65 **c. ARNg**

El ARNg proporciona el direccionamiento del sistema basado en CRISPR/Cas9. El ARNg es una fusión de dos ARN no codificantes: un ARNcr y un ARNtracr. El ARNgr puede dirigirse a cualquier secuencia de ADN deseada intercambiando la secuencia que codifica un protoespaciador de 20 pb que confiere especificidad de dirección a través del emparejamiento de bases complementarias con la diana de ADN deseada. El ARNg imita el dúplex natural ARNcr:ARNtracr implicado en el sistema efector tipo II. Este dúplex, que puede incluir, por ejemplo, un ARNcr de 42 nucleótidos y un ARNtracr de 75 nucleótidos, actúa como una guía para que Cas9 escinda el ácido nucleico diana. La "región diana", la "secuencia diana" o el "protoespaciador" como se usan indistintamente en el presente documento se refieren a la región del gen diana a la que se dirige el sistema basado en CRISPR/Cas9. El sistema basado en CRISPR/Cas9 puede incluir al menos un ARNg, en donde los ARNg se dirigen a diferentes secuencias de ADN. Las secuencias de ADN diana pueden estar superpuestas. La secuencia diana o el protoespaciador van seguidos de una secuencia PAM en el extremo 3' del protoespaciador. Los diferentes sistemas de Tipo II tienen diferentes requisitos de PAM. Por ejemplo, el sistema Tipo II de *Streptococcus pyogenes* usa una secuencia "NGG", donde "N" puede ser cualquier nucleótido.

El número de ARNg administrados a la célula puede ser al menos 1 ARNg, al menos 2 ARNg diferentes, al menos 3 ARNg diferentes, al menos 4 ARNg diferentes, al menos 5 ARNg diferentes, al menos 6 ARNg diferentes, al menos 7 ARNg diferentes, al menos 8 ARNg diferentes, al menos 9 ARNg diferentes, al menos 10 ARNg diferentes, al menos 11 ARNg diferentes, al menos 12 ARNg diferentes, al menos 13 ARNg diferentes, al menos 14 ARNg diferentes, al menos 15 ARNg diferentes, al menos 16 ARNg diferentes, al menos 17 ARNg diferentes, al menos 18 ARNg diferentes, al menos 19 ARNg diferentes, al menos 20 ARNg diferentes, al menos 25 ARNg diferentes, al menos 30 ARNg diferentes, al menos 35 ARNg diferentes, al menos 40 ARNg diferentes, al menos 45 ARNg diferentes o al menos 50 ARNg diferentes. El número de ARNg administrados a la célula puede estar entre al menos 1 ARNg y al menos 50 ARNg diferentes, al menos 1 ARNg a al menos 45 ARNg diferentes, al menos 1 ARNg a al menos 40 ARNg diferentes, al menos 1 ARNg a al menos 35 ARNg diferentes, al menos 1 ARNg a al menos 30 ARNg diferentes, al menos 1 ARNg a al menos 25 ARNg diferentes, al menos 1 ARNg a al menos 20 ARNg diferentes, al menos 1 ARNg a al menos 16 ARNg diferentes, al menos 1 ARNg a al menos 12 ARNg diferentes, al menos 1 ARNg a al menos 8 ARNg diferentes, al menos 1 ARNg a al menos 4 ARNg diferentes, al menos 4 ARNg a al menos 50 ARNg diferentes, al menos 4 ARNg diferentes a al menos 45 ARNg diferentes, al menos 4 ARNg diferentes a al menos 40 ARNg diferentes, al menos 4 ARNg diferentes a al menos 35 ARNg diferentes, al menos 4 ARNg diferentes a al menos 30 ARNg diferentes, al menos 4 ARNg diferentes a al menos 25 ARNg diferentes, al menos 4 ARNg diferentes a al menos 20 ARNg diferentes, al menos 4 ARNg diferentes a al menos 16 ARNg diferentes, al menos 4 ARNg diferentes a al menos 12 ARNg diferentes, al menos 4 ARNg diferentes a al menos 8 ARNg diferentes, al menos 8 ARNg diferentes a al menos 50 ARNg diferentes, al menos 8 ARNg diferentes a al menos 45 ARNg diferentes, al menos 8 ARNg diferentes a al menos 40 ARNg diferentes, al menos 8 ARNg diferentes a al menos 35 ARNg diferentes, al menos 8 ARNg diferentes a al menos 30 ARNg diferentes, al menos 8 ARNg diferentes a al menos 25 ARNg diferentes, 8 ARNg diferentes a al menos 20 ARNg diferentes, al menos 8 ARNg diferentes a al menos 16 ARNg diferentes u 8 ARNg diferentes a al menos 12 ARNg diferentes.

El ARNg puede comprender una secuencia polinucleotídica complementaria de la secuencia de ADN diana seguida de una secuencia PAM. El ARNg puede comprender una "G" en el extremo 5' de la secuencia polinucleotídica complementaria. El ARNg puede comprender una secuencia polinucleotídica al menos 10 pares de bases, al menos todos los pares de bases, al menos 12 pares de bases, al menos 13 pares de bases, al menos 14 pares de bases, al menos 15 pares de bases, al menos 16 pares de bases, al menos 17 pares de bases, al menos 18 pares de bases, al menos 19 pares de bases, al menos 20 pares de bases, al menos 21 pares de bases, al menos 22 pares de bases, al menos 23 pares de bases, al menos 24 pares de bases, al menos 25 pares de bases, al menos 30 pares de bases o al menos 35 pares de bases complementaria de la secuencia de ADN diana seguida de una secuencia PAM. La secuencia PAM puede ser "NGG", donde "N" puede ser cualquier nucleótido. El ARNg puede dirigirse a al menos una de la región promotora, la región potenciadora o la región transcrita del gen diana. El ARNg puede incluir una secuencia de ácido nucleico de al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563, 585-625, 462 (Fig. 40), 464 (Fig. 41) y 465 (Fig. 41).

El ARNg puede dirigirse a cualquier secuencia de ácido nucleico. La secuencia de ácido nucleico diana puede ser ADN. El ADN puede ser cualquier gen. Por ejemplo, el ARNg puede dirigirse a un gen, tales como *BRN2*, *MYT1L*, *ASCL1*, *NANOG*, *VEGFA*, *TERT*, *IL1B*, *IL1R2*, *IL1RN*, *HBG1*, *HBG2*, *MYOD1*, *OCT4* y *DMD*.

(1) Distrofina

La distrofina es una proteína citoplasmática en forma de varilla que forma parte de un complejo proteico que conecta el citoesqueleto de una fibra muscular con la matriz extracelular circundante a través de la membrana celular. La distrofina proporciona estabilidad estructural al complejo de distroglicano de la membrana celular. El gen de la distrofina tiene 2,2 megabases en el locus Xp21. La transcripción primaria mide aproximadamente 2.400 kb y el ARNm maduro aproximadamente 14 kb. 79 exones codifican la proteína que tiene más de 3500 aminoácidos. El tejido muscular esquelético normal contiene sólo pequeñas cantidades de distrofina pero su ausencia de expresión anormal conduce al desarrollo de síntomas graves e incurables. Algunas mutaciones en el gen de la distrofina conducen a la producción de distrofina defectuosa y a un fenotipo distrófico grave en los pacientes afectados. Algunas mutaciones en el gen de la distrofina conducen a una proteína distrofina parcialmente funcional y a un fenotipo

distrófico mucho más leve en los pacientes afectados.

La DMD es el resultado de mutaciones heredadas o espontáneas que provocan mutaciones sin sentido o de cambio de marco en el gen de la distrofina. Las mutaciones de origen natural y sus consecuencias se conocen relativamente bien para la DMD. Se sabe que las eliminaciones en marco que se producen en la región del exón 45-55 contenida dentro del dominio de varilla pueden producir proteínas distrofina altamente funcionales y muchos portadores son asintomáticos o muestran síntomas leves. Adicionalmente, más del 60 % de los pacientes teóricamente pueden ser tratados dirigiéndose a exones en esta región del gen de la distrofina. Se han realizado esfuerzos para restaurar el marco de lectura de distrofina alterado en pacientes con DMD omitiendo exones no esenciales durante el corte y empalme del ARNm para producir proteínas de distrofina funcionalmente eliminadas internamente. La eliminación de exones internos de distrofina conserva el marco de lectura adecuado pero provoca la distrofia muscular de Becker, menos grave.

(2) Sistema basado en CRISPR/Cas9 para dirigirse a la distrofina

En el presente documento se divulga un sistema basado en CRISPR/Cas9 específico para el gen de la distrofina. El sistema basado en CRISPR/Cas9 puede incluir Cas9 y al menos un ARNg para dirigirse al gen de la distrofina. El sistema basado en CRISPR/Cas9 puede unirse a y reconocer una región diana. Las regiones diana pueden elegirse inmediatamente en dirección 5' de posibles codones de parada fuera del marco de lectura de tal manera que las inserciones o eliminaciones durante el proceso de reparación restablezcan la conversión de marco a marco de lectura de distrofina. Las regiones diana también pueden ser sitios aceptores de corte y empalme o sitios donantes de corte y empalme, de tal manera que las inserciones o eliminaciones durante el proceso de reparación interrumpan el corte y empalme y restablezcan el marco de lectura de distrofina mediante la interrupción del sitio de corte y empalme y la exclusión de exones. Las regiones diana también pueden ser codones de parada aberrantes, de tal manera que las inserciones o las eliminaciones durante el proceso de reparación restablezcan el marco de lectura de distrofina eliminando o alterando el codón de parada.

Pueden diseñarse ARNg individuales o multiplexados para restaurar el marco de lectura de la distrofina dirigiéndose al punto de acceso mutacional en los exones 45-55 e introduciendo pequeñas inserciones y eliminaciones intraexónicas o grandes eliminaciones de uno o más exones. Después del tratamiento con Cas9 y uno o más ARNg, la expresión de distrofina puede restaurarse en las células musculares del paciente de Duchenne *in vitro*. Se detectó distrofina humana *in vivo* tras el trasplante de células de pacientes genéticamente corregidas en ratones inmunodeficientes. De manera significativa, las capacidades únicas de edición génica multiplex del sistema CRISPR/Cas9 permiten generar de manera eficiente grandes eliminaciones de esta región de punto de acceso mutacional que puede corregir hasta el 62 % de las mutaciones de los pacientes mediante enfoques de edición génica universales o específicos del paciente.

El sistema basado en CRISPR/Cas9 puede usar ARNg de diferentes secuencias y longitudes. Pueden encontrarse ejemplos de ARNg en la Tabla 6. El sistema basado en CRISPR/Cas9 puede dirigirse a una secuencia de ácido nucleico de SEQ ID NO: 65-144 o un complemento de la misma. El ARNg puede incluir una secuencia nucleotídica seleccionada del grupo que consiste en SEQ ID NO: 65-144 o un complemento de la misma. Por ejemplo, los sistemas basados en CRISPR/Cas9 divulgados fueron diseñados para mediar en la edición génica altamente eficiente en el exón 51 del gen de la distrofina. Estos sistemas basados en CRISPR/Cas9 restauraron la expresión de la proteína distrofina en células de pacientes con DMD.

(a) Exones 51 y 45-55

El exón 51 suele estar adyacente a eliminaciones que alteran el marco en la DMD. La eliminación del exón 51 del transcripto de distrofina mediante la omisión del exón puede usarse para tratar aproximadamente el 15 % de todos los pacientes con DMD. Esta clase de mutaciones DMD es ideal para la corrección permanente mediante edición genómica basada en NHEJ y HDR. Los sistemas basados en CRISPR/Cas9 descritos en el presente documento se han desarrollado para la modificación dirigida del exón 51 en el gen de la distrofina humana. Estos sistemas basados en CRISPR/Cas9 se transfecaron en células DMD humanas y mediaron en la modificación génica eficiente y la conversión al marco de lectura correcto. La restauración de proteínas fue concomitante con la restauración del marco y se detectó en una población masiva de células tratadas con el sistema basado en CRISPR/Cas9. De forma similar, la eliminación de los exones 45-55 del transcripto de distrofina puede usarse para tratar aproximadamente el 62 % de todos los pacientes con DMD.

(3) Construcciones AAV/CRISPR

Puede usarse AAV para suministrar CRISPR usando diversas configuraciones de construcción (véase la Fig. 39). Por ejemplo, AAV puede suministrar casetes de expresión de Cas9 y ARNg en vectores separados. Como alternativa, si las pequeñas proteínas Cas9, derivadas de especies como *Staphylococcus aureus* o *Neisseria meningitidis*, se usan entonces tanto Cas9 como hasta dos casetes de expresión de ARNg pueden combinarse en un único vector AAV dentro del límite de empaquetamiento de 4,7 kb (véase la Fig. 39).

5. Sistema multiplex basado en CRISPR/Cas9

La presente divulgación se dirige a un sistema múltiplex basado en CRISPR/Cas9 que incluye un sistema basado en CRISPR/asociado a CRISPR (Cas) 9, tal como Cas9 o dCas9, y múltiples ARNg para dirigirse a uno o más genes endógenos. Esta plataforma usa un conveniente método de clonación Golden Gate para incorporar rápidamente hasta cuatro casetes de expresión de ARNg independientes en un único vector lentivírico. Cada ARNg se expresó de manera eficiente y podría mediar en la edición génica múltiplex en diversos loci en líneas celulares humanas primarias e inmortalizadas. Se demostró que la activación transcripcional transitoria en líneas celulares que expresan establemente dCas9-VP64 puede ajustarse mediante activación sinérgica con uno a cuatro ARNg. Adicionalmente, el vector lentivírico único puede inducir la expresión génica endógena sostenida y a largo plazo en células humanas primarias e inmortalizadas. Este sistema permite el ensamblaje rápido de un único vector lentivírico que permite la edición o la activación múltiple eficiente de genes en líneas celulares modelo y primarias.

El sistema multiplex basado en CRISPR/Cas9 proporciona potencia de activación transcripcional e inducción sintonizable de activación transcripcional. Fácilmente generado por el ensamblaje Golden Gate, el vector final expresa un Cas9 o dCas9-VP64 constitutivo además de uno, dos, tres o cuatro ARNg expresados a partir de promotores independientes. Cada promotor es capaz de expresar eficientemente ARNg que dirigen niveles similares de actividad de la nucleasa Cas9. Adicionalmente, el suministro lentivírico de un único vector que expresa Cas9 y cuatro ARNg dirigidos a loci independientes dio como resultado la edición múltiple simultánea de genes de los cuatro loci. La activación transcripcional sintonizable en dos genes endógenos en contextos transitorios y estables se logró mediante el suministro lentivírico de Cas9 con o sin ARNg. Se logra una activación génica altamente eficiente y a largo plazo en células humanas primarias. Por lo tanto, este sistema es un método atractivo y eficiente para generar edición genética múltiple y activación transcripcional a largo plazo en células humanas.

El sistema multiplex basado en CRISPR/Cas9 permite la edición eficiente de genes multiplex para inactivar simultáneamente múltiples genes. El sistema CRISPR/Cas9 puede dirigirse simultáneamente a múltiples loci genómicos distintos al coexpresar una única proteína Cas9 con dos o más ARNg, lo que hace que este sistema sea especialmente adecuado para aplicaciones de activación sinérgica o edición génica multiplex. El sistema CRISPR/Cas9 acelera enormemente el proceso de direccionamiento molecular a nuevos sitios simplemente modificando la molécula de ARNg expresada. El vector lentivírico individual puede combinarse con métodos para lograr el control inducible de estos componentes, ya sea por regulación química u optogenética, para facilitar la investigación de la dinámica de la regulación genética tanto en el tiempo como en el espacio.

Los sistemas multiplex basados en CRISPR/Cas9 pueden activar transcripcionalmente dos o más genes endógenos. Los sistemas múltiples basados en CRISPR/Cas9 pueden reprimir transcripcionalmente dos o más genes endógenos. Por ejemplo, al menos dos genes endógenos, al menos tres genes endógenos, al menos cuatro genes endógenos, al menos cinco genes endógenos o al menos diez genes endógenos pueden estar activados o reprimidos por el sistema multiplex basado en CRISPR/Cas9. Entre dos y quince genes, entre dos y diez genes, entre dos y cinco genes, entre cinco y quince genes o entre cinco y diez genes pueden estar activados o reprimidos por el sistema multiplex basado en CRISPR/Cas9.

(1) Vector lentivírico modificado

El sistema multiplex basado en CRISPR/Cas9 incluye un vector lentivírico modificado. El vector lentivírico modificado incluye una primera secuencia polinucleotídica que codifica una proteína de fusión y una segunda secuencia polinucleotídica que codifica al menos un ARNg. La proteína de fusión puede ser la proteína de fusión del sistema basado en CRISPR/Cas9, como se describió anteriormente. La primera secuencia polinucleotídica puede estar unida operativamente a un promotor. El promotor puede ser un promotor constitutivo, un promotor inducible, un promotor reprimible o un promotor regulable.

La segunda secuencia polinucleotídica codifica al menos 1 ARNg. Por ejemplo, la segunda secuencia polinucleotídica puede codificar al menos 1 ARNg, al menos 2 ARNg, al menos 3 ARNg, al menos 4 ARNg, al menos 5 ARNg, al menos 6 ARNg, al menos 7 ARNg, al menos 8 ARNg, al menos 9 ARNg, al menos 10 ARNg, al menos 11 ARNg, al menos 12 ARNg, al menos 13 ARNg, al menos 14 ARNg, al menos 15 ARNg, al menos 16 ARNg, al menos 17 ARNg, al menos 18 ARNg, al menos 19 ARNg, al menos 20 ARNg, al menos 25 ARNg, al menos 30 ARNg, al menos 35 ARNg, al menos 40 ARNg, al menos 45 ARNg o al menos 50 ARNg. La segunda secuencia polinucleotídica puede codificar entre 1 ARNg y 50 ARNg, entre 1 ARNg y 45 ARNg, entre 1 ARNg y 40 ARNg, entre 1 ARNg y 35 ARNg, entre 1 ARNg y 30 ARNg, entre 1 ARNg y 25 ARNg diferentes, entre 1 ARNg y 20 ARNg, entre 1 ARNg y 16 ARNg, entre 1 ARNg y 8 ARNg diferentes, entre 4 ARNg diferentes y 50 ARNg diferentes, entre 4 ARNg diferentes y 45 ARNg diferentes, entre 4 ARNg diferentes y 40 ARNg diferentes, entre 4 ARNg diferentes y 35 ARNg diferentes, entre 4 ARNg diferentes y 30 ARNg diferentes, entre 4 ARNg diferentes y 25 ARNg diferentes, entre 4 ARNg diferentes y 20 ARNg diferentes, entre 4 ARNg diferentes y 16 ARNg diferentes, entre 4 ARNg diferentes y 8 ARNg diferentes, entre 8 ARNg diferentes y 50 ARNg diferentes, entre 8 ARNg diferentes y 45 ARNg diferentes, entre 8 ARNg diferentes y 40 ARNg diferentes, entre 8 ARNg diferentes y 35 ARNg diferentes, entre 8 ARNg diferentes y 30 ARNg diferentes, entre 8 ARNg diferentes y 25 ARNg diferentes, entre

8 ARN_{gm} diferentes y 20 ARN_{gm} diferentes, entre 8 ARN_{gm} diferentes y 16 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 50 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 45 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 40 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 35 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 30 ARN_{gm} diferentes, entre 16 ARN_{gm} diferentes y 25 ARN_{gm} diferentes o entre 16 ARN_{gm} diferentes y 20 ARN_{gm} diferentes.

5 Cada una de las secuencias de polinucleótidos que codifican los diferentes ARN_{gm} pueden estar unidas operativamente a un promotor. Los promotores que están unidos operativamente a los diferentes ARN_{gm} pueden ser el mismo promotor. Los promotores que están unidos operativamente a los diferentes ARN_{gm} pueden ser promotores diferentes. El promotor puede ser un promotor constitutivo, un promotor inducible, un promotor reprimible o un promotor regulable.

10 El al menos un ARN_{gm} puede unirse a un gen o loci diana. Si se incluye más de un ARN_{gm}, cada uno de los ARN_{gm} se une a una región diana diferente dentro de un loci diana o cada uno de los ARN_{gm} se une a una región diana diferente dentro de diferentes loci genéticos. La proteína de fusión puede incluir la proteína Cas9 o la proteína iCas9-VP64. La proteína de fusión puede incluir un dominio VP64, un dominio p300 o un dominio KRAB.

15 **6. Nucleasas específicas de sitio**

La composición, como se ha descrito anteriormente, incluye una secuencia nucleotídica que codifica una nucleasa específica de sitio que se une y escinde una región diana. La nucleasa específica de sitio puede diseñarse por ingeniería genética. Por ejemplo, una nucleasa específica de sitio diseñada puede ser un sistema basado en CRISPR/Cas9, una ZFN o una TALEN. La nucleasa específica de sitio puede unirse y escindir un gen o locus en el genoma de una célula en el músculo esquelético o el músculo cardíaco. Por ejemplo, el gen o locus puede ser el locus Rosa26 o el gen de la distrofina.

25 **a. Sistema basado en CRISPR/Cas9**

El sistema basado en CRISPR/Cas9, como se ha descrito anteriormente, puede usarse para introducir roturas bicatenarias específicas de sitio en loci genómicos específicos.

30 **b. Nucleasas con dedos de cinc (ZFN)**

La nucleasa específica de sitio puede ser una ZFN. Un solo dedo de cinc contiene aproximadamente 30 aminoácidos y el dominio funciona uniendo 3 pares de bases consecutivos de ADN mediante interacciones de una única cadena lateral de aminoácido por par de bases. La estructura modular del motivo de los dedos de cinc permite la conjunción de varios dominios en serie, permitiendo el reconocimiento y el direccionamiento de secuencias extendidas en múltiples de 3 nucleótidos. Estos dominios de unión al ADN específicos pueden combinarse con un dominio de nucleasa, tales como FokI, para generar una nucleasa específica de sitio, llamada "nucleasa de dedos de cinc" (ZFN) que puede usarse para introducir roturas bicatenarias específicas de un sitio en loci genómicos específicos. Esta escisión del ADN estimula la maquinaria natural de reparación del ADN, lo que lleva a una de las dos posibles vías de reparación, NHEJ y HDR. Por ejemplo, la ZFN puede dirigirse al locus Rosa26 (Pérez-Piñera *et al.* Nucleic Acids Research (2012) 40:3741-3752) o un gen de distrofina. Los ejemplos de las ZFN se muestran en la Tabla 1 y las Figs. 35-38. En la Tabla 1, las hélices de reconocimiento de ADN están subrayadas y "Fok ELD-S" y "Fok KKR-S" se refieren al dominio de nucleasa FokI que está fusionado a los dominios de unión al ADN de la proteína con dedos de cinc. En las Fig. 35-38 la secuencia de ADN diana en los sitios diana (es decir, en SEQ ID NO: 442, 445, 448 y 453) y las hélices de reconocimiento de ADN en las secuencias de aminoácidos de ZFN (es decir, en SEQ ID NO: 443, 444, 446, 447, 449-452, 454 y 455) están subrayadas, respectivamente.

Tabla 1 Secuencias de aminoácidos completas de ZFN identificadas.

Secuencia de aminoácidos completa de ZFN B izquierda Fok ELD-S (SEQ ID NO:438)
MDYKDHDGDYKDHDIDYKDDDKMAPKKRKVGRLEPGEKPYKCPECGKSFSRKDALRGH QRTHTGEKPYKCPECGKSFS <u>HRTTLTNH</u> QRTHTGEKPYKCPECGKSFS <u>ORNALAGH</u> QRTGH TGEKPYKCPECGKSFS <u>HKNALONH</u> QRTHTGEKPYKCPECGKSFS <u>DPGHLVRH</u> QRTHTGEK PYKCPECGKSFS <u>TSGNLVRH</u> QRTHTGAAARALVKSELEEKSELRHKLKYVPHEYIELIE IARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGIVDTKAYSG GYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNHITNCNGAVLSVEELLIGEMIKAGTLTLEEVRKFNNGEINF*

Secuencia de aminoácidos completa de ZFN B derecha Fok KKR-S (SEQ ID NO:439)

<pre> MDYKDHDGDYKDHDIDYKDDDKMAPKKRKVGRLEPGEKPYKCPECGKSFSOORSLVGH QRTHTGEKPYKCPECGKSFS<u>DKKD</u>LTRHQRTHTGEKPYKCPECGKSFS<u>TSGH</u>LVRHQRT TGEKPYKCPECGKSFS<u>QRAHLERH</u>QRTHTGEKPYKCPECGKSFS<u>TSGS</u>LV<u>RH</u>QRTHTGEK PYKCPECGKSFS<u>TSGN</u>LVHQRTHTGAAARALVKSELEEKSELRHKLKYVPHEYIELIE IARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSG GYNLPIGQADEMQRYYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNRKTNNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF* </pre> <p>Secuencia de aminoácidos completa de ZFN J izquierda Fok KKR-S (SEQ ID NO:440)</p> <pre> MDYKDHDGDYKDHDIDYKDDDKMAPKKRKVGRLEPGERPFQCRICMRNFSS<u>KOALAVH</u> TRTHTGEKPFQCRICMRNF<u>SQSTTL</u>KRHLRHTGEKPFQCRICMRNF<u>SDHLSL</u>HLKTH LRGSQLVKSELEEKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYR GEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGGYNLPIGQADEMQRYYVKENQTRNHI NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLN<u>RKTNCNGAVLSVEELLIGGEMIK</u> AGTLTLEEVRRKFNNGEINF* </pre> <p>Secuencia de aminoácidos completa de ZFN B derecha Fok ELD-S (SEQ ID NO:441)</p> <pre> MDYKDHDGDYKDHDIDYKDDDKMAPKKRKVGRLEPGERPFQCRICMRNF<u>SRRAHLQN</u>H TRTHTGEKPFQCRICMRNF<u>SQSTTL</u>KRHLRHTGEKPFQCRICMRNF<u>SDGGHL</u>TRHLKTH LRGSQLVKSELEEKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYR GEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGGYNLPIGQADEMERYVEENQTRDKHL NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLN<u>HITNCNGAVLSVEELLIGGEMIK</u> AGTLTLEEVRRKFNNGEINF* </pre>

c. Nucleasas efectoras TAL (TALEN)

- Las TALEN pueden usarse para introducir roturas bicatenarias específicas de sitio en loci genómicos específicos. Las 5 roturas bicatenarias específicas del sitio se crean cuando dos TALEN independientes se unen a secuencias de ADN cercanas, permitiendo de esta manera la dimerización de FokI y la escisión del ADN diana. Los TALEN tienen una edición avanzada del genoma debido a su alta tasa de modificación génica exitosa y eficiente. Esta escisión del ADN puede estimular la maquinaria natural de reparación del ADN, conduciendo a una de dos posibles vías de reparación: 10 reparación dirigida por homología (HDR) o la vía de unión de extremos no homólogos (NHEJ). Las TALEN pueden diseñarse para dirigirse a cualquier gen implicado en una enfermedad genética.

Las TALEN pueden incluir una nucleasa y un dominio de unión al ADN de TALE que se une al gen diana en una región diana de TALEN. El gen diana puede tener una mutación tal como una mutación de cambio de marco o una mutación sin sentido. Si el gen diana tiene una mutación que provoca un codón de parada prematuro, la TALEN puede diseñarse 15 para reconocer y unirse a una secuencia nucleotídica en dirección 5' o en dirección 3' del codón de parada prematuro. Una "región diana de TALEN" incluye las regiones de unión para dos TALEN y la región espaciadora, que se produce entre las regiones de unión. Los dos TALEN se unen a diferentes regiones de unión dentro de la región diana de TALEN, después de lo cual se escinde la región diana de TALEN. Algunos ejemplos de TALEN se describen en la Solicitud de Patente Internacional N.º PCT/US2013/038536.

20

7. Activadores transcripcionales

- La composición, como se ha descrito anteriormente, incluye una secuencia nucleotídica que codifica un activador transcripcional que activa un gen diana. El activador transcripcional puede modificarse por ingeniería genética. Por ejemplo, un activador transcripcional modificado por ingeniería genética puede ser un sistema basado en 25 CRISPR/Cas9, una proteína de fusión con dedos de cinc o una proteína de fusión TALE.

a. Sistema basado en CRISPR/Cas9

- El sistema basado en CRISPR/Cas9, como se ha descrito anteriormente, puede usarse para activar la transcripción 30 de un gen diana con ARN. El sistema basado en CRISPR/Cas9 puede incluir una proteína de fusión, como se ha descrito anteriormente, en donde el segundo dominio polipeptídico tiene actividad de activación de la transcripción o actividad de modificación de histonas. Por ejemplo, el segundo dominio polipeptídico puede incluir VP64 o p300.

b. Proteínas de fusión con dedos de cinc

5 El activador transcripcional puede ser una proteína de fusión con dedos de cinc. El dedo de cinc direccionalizado a los dominios de unión al ADN, como se ha descrito anteriormente, puede combinarse con un dominio que tiene actividad de activación de la transcripción o actividad de modificación de histonas. Por ejemplo, el dominio puede incluir VP64 o p300.

c. Proteínas de fusión TALE

10 Pueden usarse proteínas de fusión TALE para activar la transcripción de un gen diana. La proteína de fusión TALE puede incluir un dominio de unión al ADN de TALE y un dominio que tiene actividad de activación de la transcripción o actividad de modificación de histonas. Por ejemplo, el dominio puede incluir VP64 o p300.

15 8. Composiciones

La presente divulgación se dirige a una composición para alterar la expresión génica y modificar por ingeniería genética o alterar el ADN genómico en una célula o sujeto. La composición también puede incluir un sistema de suministro vírico.

a. Composiciones para la edición genómica en el músculo

20 La presente divulgación se dirige a una composición para editar el genoma de un gen diana en el músculo esquelético o el músculo cardíaco de un sujeto. La composición incluye un vector AAV modificado y una secuencia nucleotídica que codifica una nucleasa específica de sitio. La composición suministra formas activas de nucleasas específicas de sitio al músculo esquelético o al músculo cardíaco. La composición puede comprender además un ADN donante o un transgén. Estas composiciones pueden usarse en edición genómica, ingeniería genómica y corrección o reducción de los efectos de mutaciones en genes implicados en enfermedades genéticas y/u otras afecciones del músculo esquelético o cardíaco.

25 30 El gen diana puede estar implicado en la diferenciación de una célula o en cualquier otro proceso donde la activación, la represión o la alteración de un gen puede desearse, o puede tener una mutación tal como una eliminación, una mutación de desplazamiento de marco o una mutación sin sentido. Si el gen diana tiene una mutación que provoca un codón de parada prematuro, un sitio aceptor de corte y empalme aberrante o un sitio donante de corte y empalme aberrante, las nucleasas específicas de sitio pueden diseñarse para reconocer y unirse a una secuencia nucleotídica en dirección 5' o en dirección 3' del codón de parada prematuro, el sitio aceptor de corte y empalme aberrante o el sitio donante de corte y empalme aberrante. Las nucleasas específicas de sitio también pueden usarse para alterar el corte y empalme normal de genes dirigiéndose a los aceptores y donantes del corte y empalme para inducir la omisión de codones de parada prematuros o restaurar un marco de lectura alterado. Las nucleasas específicas de sitio pueden 35 40 mediar o no cambios inespecífico en regiones del genoma que codifican proteínas.

b. Vectores de virus adenoasociados

45 La composición, como se ha descrito anteriormente, incluye un vector de virus adenoasociado (AAV) modificado. El vector AAV modificado puede haber mejorado el tropismo del tejido del músculo cardíaco y esquelético. El vector AAV modificado puede ser capaz de suministrar y expresar la nucleasa específica de sitio en la célula de un mamífero. Por ejemplo, el vector AAV modificado puede ser un vector AAV-SASTG (Piacentino *et al.* (2012) Human Gene Therapy 23:635-646). El vector AAV modificado puede suministrar nucleasas al músculo esquelético y cardíaco *in vivo*. El vector AAV modificado puede basarse en uno o más de varios tipos de cápside, incluyendo AAV1, AAV2, AAV5, AAV6, 50 AAV8 y AAV9. El vector de AAV modificado puede basarse en el pseudotipo de AAV2 con cápsides de AAV con tropismo muscular alternativo, tales como AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 y vectores AAV/SASTG que transducen eficientemente el músculo esquelético o el músculo cardíaco mediante suministro sistémica y local (Seto *et al.* Current Gene Therapy (2012) 12:139-151).

55 c. Sistema basado en CRISPR/Cas9

La presente divulgación también proporciona sistemas dirigidos a ADN o composiciones de al menos un sistema basado en CRISPR/Cas9, como se describió anteriormente. Estas composiciones pueden usarse en edición genómica, ingeniería genómica y corrección o reducción de los efectos de mutaciones en genes implicados en enfermedades genéticas. La composición incluye un sistema basado en CRISPR/Cas9 que incluye una proteína Cas9 o una proteína de fusión Cas9, como se describió anteriormente. El sistema basado en CRISPR/Cas9 también puede incluir al menos un ARNg, como se describió anteriormente.

65 d. Sistema multiplex basado en CRISPR/Cas9

La presente divulgación también proporciona sistemas multiplex basados en CRISPR/Cas9, como se describió

anteriormente. Estas composiciones pueden usarse en edición genómica, ingeniería genómica y corrección o reducción de los efectos de mutaciones en genes implicados en enfermedades genéticas. Estas composiciones pueden usarse para dirigirse a más de un gen. La composición incluye un vector lentivírico modificado que comprende un sistema basado en CRISPR/Cas9 que incluye una proteína Cas9 o una proteína de fusión Cas9, como se describió anteriormente y más de un ARNg, como se describió anteriormente.

9. Métodos de uso

Las aplicaciones potenciales de las composiciones son diversas en muchas áreas de la ciencia y la biotecnología. Las 10 composiciones divulgadas pueden usarse para reparar mutaciones genéticas que provocan enfermedades. Las composiciones divulgadas pueden usarse para alterar genes de tal manera que la alteración genética conduce a aumentos en la regeneración muscular o la fuerza muscular, o disminuciones en el envejecimiento muscular. Las composiciones divulgadas pueden usarse para introducir genes terapéuticos que se expresarán sistémicamente a partir del músculo esquelético o del músculo cardíaco, tales como factores de coagulación o anticuerpos monoclonales. 15 Las composiciones divulgadas pueden usarse para modular la expresión de genes de mamíferos. Las composiciones divulgadas pueden usarse para transdiferenciar o inducir la diferenciación de una célula o corregir un gen mutante en una célula. Se proporcionan ejemplos de activación de genes relacionados con la terapia celular y génica, reprogramación genética y medicina regenerativa. Pueden usarse activadores transcripcionales guiados por ARN para reprogramar la especificación del linaje celular. La activación de genes endógenos que codifican los reguladores clave 20 del destino celular, en lugar de una sobreexpresión forzada de estos factores, potencialmente puede conducir a métodos rápidos, eficientes, estables o específicos para la reprogramación genética, la transdiferenciación y/o la diferenciación inducida.

10. Uso en métodos de edición genómica en músculos

La presente divulgación es para su uso en un método de edición genómica en un músculo esquelético o músculo 25 cardíaco de un sujeto. El uso comprende administrar al músculo esquelético o al músculo cardíaco del sujeto la composición para la edición genómica en el músculo esquelético o al músculo cardíaco, como se describió anteriormente. La edición genómica puede incluir corregir un gen mutante o insertar un transgén. Corregir el gen mutante puede incluir eliminar, reorganizar o reemplazar el gen mutante. Corregir el gen mutante puede incluir NHEJ 30 o HDR mediado por nucleasa.

11. Métodos de uso del sistema basado en CRISPR/Cas9

Las posibles aplicaciones del sistema basado en CRISPR/Cas9 son diversas en muchas áreas de la ciencia y la 35 biotecnología. Los sistemas basados en CRISPR/Cas9 divulgados pueden usarse para modular la expresión de genes de mamíferos. Los sistemas basados en CRISPR/Cas9 divulgados pueden usarse para transdiferenciar o inducir la diferenciación de una célula o corregir un gen mutante en una célula. Se proporcionan ejemplos de activación de genes relacionados con la terapia celular y génica, reprogramación genética y medicina regenerativa. Pueden usarse 40 activadores transcripcionales guiados por ARN para reprogramar la especificación del linaje celular. Aunque la reprogramación fue incompleta e ineficiente en estos experimentos, hay muchas maneras en que este método podría mejorarse, incluyendo transfecciones repetidas de combinaciones iCas9-VP64/ARNg, expresión estable de estos factores y dirigida a múltiples genes, tales como Bm2 y Mytll además de Ascl1 para la transdiferenciación en un fenotipo neuronal. La activación de genes endógenos que codifican los reguladores clave del destino celular, en lugar de una 45 sobreexpresión forzada de estos factores, potencialmente puede conducir a métodos rápidos, eficientes, estables o específicos para la reprogramación genética y la transdiferenciación o la diferenciación inducida. Finalmente, las fusiones de Cas9 con otros dominios, incluyendo dominios represivos y modificadores epigenéticos, podría proporcionar una mayor diversidad de reguladores transcripcionales guiados por ARN para complementar otras 50 herramientas basadas en ARN para la ingeniería de células de mamíferos.

a. Uso en métodos de activación de la expresión génica

La presente divulgación proporciona un mecanismo para activar la expresión de genes endógenos, tales como genes 55 de mamíferos, basados en dirigir un activador transcripcional a promotores a través de ARN usando un sistema basado en CRISPR/Cas9, como se describió anteriormente. Esto es fundamentalmente diferente de los métodos descritos anteriormente basados en la ingeniería de proteínas de unión al ADN de secuencias específicas y puede brindar oportunidades para la regulación genética específica. Debido a que la generación de plásmidos de expresión de ARNg simplemente implica la síntesis de dos oligonucleótidos cortos personalizados y una etapa de clonación, es posible 60 generar muchos activadores de genes nuevos de forma rápida y económica. Los ARNg también pueden transfectarse directamente a las células después de la transcripción *in vitro*. Se mostraron múltiples ARNg dirigidos a promotores individuales, pero también podría ser posible dirigirse simultáneamente a múltiples promotores. El reconocimiento de sitios diana genómicos con ARN, en lugar de proteínas, también puede eludir las limitaciones de dirigirse a sitios modificados epigenéticamente, tales como el ADN metilado.

A diferencia de los métodos actuales basados en la ingeniería de proteínas de unión al ADN, Cas9 fusionado a un 65 dominio de activación transcripcional puede direccionarse por combinaciones de moléculas de ARN guía para inducir

la expresión de genes humanos endógenos. Este enfoque sencillo y versátil para la activación génica dirigida evita la necesidad de diseñar nuevas proteínas y permite una regulación genética sintética generalizada.

5 El uso puede incluir la administración a una célula o sujeto de un sistema basado en CRISPR/Cas9, un polinucleótido o vector que codifica dicho sistema basado en CRISPR/Cas9 o sistemas de direccionamiento de ADN o composiciones de al menos un sistema basado en CRISPR/Cas9, como se describió anteriormente. El uso puede incluir la administración de un sistema basado en CRISPR/Cas9, tal como administrar una proteína de fusión Cas9 que contiene un dominio de activación de la transcripción o una secuencia nucleotídica que codifica dicha proteína de fusión Cas9. La proteína de fusión Cas9 puede incluir un dominio de activación de la transcripción tal como una proteína VP16 o un coactivador de la transcripción tal como una proteína p300.

10

(1) dCas9-VP16

15 La proteína de fusión Cas9 puede incluir un dominio de activación de la transcripción tal como la proteína aVP16. El dominio de activación de la transcripción puede contener al menos 1 proteína VP16, al menos 2 proteínas VP16, al menos 3 proteínas VP16, al menos 4 proteínas VP16 (es decir, un dominio activador de VP64), al menos 5 proteínas VP16, al menos 6 proteínas VP16, al menos 6 proteínas VP16 o al menos 10 proteínas VP16. La proteína Cas9 puede ser una proteína Cas9 en la que la actividad nucleasa está inactivada. Por ejemplo, la proteína Cas9 en la proteína de fusión puede ser iCas9 (aminoácidos 36-1403 de SEQ ID NO: 1), que incluye las sustituciones de aminoácidos de D10A y H840A. La proteína de fusión Cas9 puede ser iCas9-VP64.

20

(2) dCas9-p300

25 La proteína de fusión Cas9 puede incluir un dominio de coactivación de la transcripción tal como la proteína p300. La proteína p300 (también conocida como proteína p300 de unión a EP300 o E1A) está codificada por gen el *EP300* y regula la actividad de muchos genes en los tejidos de todo el cuerpo. La proteína p300 desempeña un papel en la regulación del crecimiento y la división celular, impulsando a las células a madurar y asumir funciones especializadas (diferenciarse) y prevenir el crecimiento de tumores cancerosos. La proteína p300 activa la transcripción conectando factores de transcripción con un complejo de proteínas que llevan a cabo la transcripción en el núcleo de la célula. La interacción de p300 con los factores de transcripción está gestionada por uno o más de los dominios de p300: el dominio de interacción del receptor nuclear (RID), el dominio de interacción CREB y MYB (KIX), las regiones de cisteína/histidina (TAZ1/CH1 y TAZ2/CH3) y el dominio de unión a la respuesta al interferón (IBID). Los últimos cuatro dominios, KJX, TAZ1, TAZ2 e IBID of p300, cada uno se une estrechamente a una secuencia que abarca ambos dominios de transactivación 9aaTAD del factor de transcripción p53. La proteína funciona como histona acetiltransferasa que regula la transcripción mediante la remodelación de la cromatina y es importante en los procesos de proliferación y diferenciación celular. Media la regulación del gen AMPc uniéndose específicamente a la proteína CREB fosforilada.

30

35

40 La proteína p300 puede activar Madres contra el homólogo decapentapléjico 7, MAF, TSG101, Receptor alfa activado por proliferador de peroxisomas, NPAS2, PAX6, DDX5, MYBL2, Madres contra el homólogo 1 decapentapléjico, Madres contra el homólogo 2 decapentapléjico, Factor 1 de unión al potenciador linfoide, SNIP1, TRERF1, STAT3, EID1, Receptor alfa huérfano relacionado con RAR, ELK1, HIF1A, ING5, Receptor gamma activado por proliferador de peroxisomas, SSI8, TCF3, Zif268, Receptor de estrógeno alfa, GPS2, MyoD, YY1, ING4, PR0X1, CITED1, HNF1A, MEF2C, MEF2D, MAML1, Factor de transcripción de torsión, PTMA, 1RF2, DTX1, Endonucleasa 1 específica de estructura de solapa, Factor potenciador específico de miocitos 2A, CDX2, BRCA1, HNRPU, STAT6, CITED2, RELA, TGS1, CEBPB, Mdm2, NC0A6, NFATC2, Receptor alfa de la hormona tiroidea, BCL3, TFAP2A, PCNA, P53 y TAL1.

45

50 El dominio de coactivación de la transcripción puede incluir una proteína p300 humana o un fragmento de la misma. El dominio de coactivación de la transcripción puede incluir una proteína p300 humana de tipo silvestre o una proteína p300 humana mutante o fragmentos de las mismas. El dominio de coactivación de la transcripción puede incluir el dominio central de lisina-acetiltranserasa de la proteína p300 humana, es decir, el núcleo p300 HAT (también conocido como "núcleo TS de p300"; véase la Fig. 58). La proteína Cas9 puede ser una proteína Cas9 en la que la actividad nucleasa está inactivada. Por ejemplo, la proteína Cas9 en la proteína de fusión puede ser iCas9 (aminoácidos 36-1403 de SEQ ID NO: 1), que incluye las sustituciones de aminoácidos de D10A y H840A. La proteína de fusión Cas9 puede ser iCas9-núcleo TS p300.

55

(3) ARNg

60 El método también puede incluir administrar a una célula o someter a un sistema basado en CRISPR/Cas9 al menos un ARNg, en donde los ARNg se dirigen a diferentes secuencias de ADN. Las secuencias de ADN diana pueden estar superpuestas. El número de ARNg administrados a la célula puede ser al menos 1 ARNg, al menos 2 ARNg diferentes, al menos 3 ARNg diferentes, al menos 4 ARNg diferentes, al menos 5 ARNg diferentes, al menos 6 ARNg diferentes, al menos 7 ARNg diferentes, al menos 8 ARNg diferentes, al menos 9 ARNg diferentes, al menos 10 ARNg diferentes, al menos 11 ARNg diferentes, al menos 12 ARNg diferentes, al menos 13 ARNg diferentes, al menos 14 ARNg diferentes, al menos 15 ARNg diferentes, al menos 16 ARNg diferentes, al menos 17 ARNg diferentes, al menos 18 ARNg diferentes, al menos 19 ARNg diferentes, al menos 20 ARNg diferentes, al menos 25 ARNg diferentes, al

65

menos 30 ARNg diferentes, al menos 35 ARNg diferentes, al menos 40 ARNg diferentes, al menos 45 ARNg diferentes o al menos 50 ARNg diferentes. El número de ARNg administrados a la célula puede estar entre al menos 1 ARNg y al menos 50 ARNg diferentes, al menos 1 ARNg a al menos 45 ARNg diferentes, al menos 1 ARNg a al menos 40 ARNg diferentes, al menos 1 ARNg a al menos 35 ARNg diferentes, al menos 1 ARNg a al menos 30 ARNg diferentes, al menos 1 ARNg a al menos 25 ARNg diferentes, al menos 1 ARNg a al menos 20 ARNg diferentes, al menos 1 ARNg a al menos 16 ARNg diferentes, al menos 1 ARNg a al menos 12 ARNg diferentes, al menos 1 ARNg a al menos 8 ARNg diferentes, al menos 1 ARNg a al menos 4 ARNg diferentes, al menos 4 ARNg a al menos 50 ARNg diferentes, al menos 4 ARNg diferentes a al menos 45 ARNg diferentes, al menos 4 ARNg diferentes a al menos 40 ARNg diferentes, al menos 4 ARNg diferentes a al menos 35 ARNg diferentes, al menos 4 ARNg diferentes a al menos 30 ARNg diferentes, al menos 4 ARNg diferentes a al menos 25 ARNg diferentes, al menos 4 ARNg diferentes a al menos 20 ARNg diferentes, al menos 4 ARNg diferentes a al menos 16 ARNg diferentes, al menos 4 ARNg diferentes a al menos 12 ARNg diferentes, al menos 4 ARNg diferentes a al menos 8 ARNg diferentes, al menos 8 ARNg diferentes a al menos 50 ARNg diferentes, al menos 8 ARNg diferentes a al menos 45 ARNg diferentes, al menos 8 ARNg diferentes a al menos 40 ARNg diferentes, al menos 8 ARNg diferentes a al menos 35 ARNg diferentes, 8 ARNg diferentes a al menos 30 ARNg diferentes, al menos 8 ARNg diferentes a al menos 25 ARNg diferentes, 8 ARNg diferentes a al menos 20 ARNg diferentes, al menos 8 ARNg diferentes a al menos 16 ARNg diferentes, 8 ARNg diferentes a al menos 12 ARNg diferentes, al menos 8 ARNg diferentes a al menos 8 ARNg diferentes,

El ARNg puede comprender una secuencia polinucleotídica complementaria de la secuencia de ADN diana seguida de NGG. El ARNg puede comprender una "G" en el extremo 5' de la secuencia polinucleotídica complementaria. El ARNg puede comprender una secuencia polinucleotídica al menos 10 pares de bases, al menos 11 pares de bases, al menos 12 pares de bases, al menos 13 pares de bases, al menos 14 pares de bases, al menos 15 pares de bases, al menos 16 pares de bases, al menos 17 pares de bases, al menos 18 pares de bases, al menos 19 pares de bases, al menos 20 pares de bases, al menos 21 pares de bases, al menos 22 pares de bases, al menos 23 pares de bases, al menos 24 pares de bases, al menos 25 pares de bases, al menos 30 pares de bases o al menos 35 pares de bases complementaria de la secuencia de ADN diana seguida de NGG. El ARNg puede dirigirse a al menos una de la región promotora, la región potenciadora o la región transcrita del gen diana. El ARNg puede incluir una secuencia de ácido nucleico de al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625.

b. Uso en métodos de represión de la expresión génica

La presente divulgación proporciona un mecanismo para reprimir la expresión de genes endógenos, tales como genes de mamíferos, basado en dirigirse a elementos reguladores genómicos, tales como potenciadores distales, a través de ARN usando un sistema basado en CRISPR/Cas9, como se describió anteriormente. La proteína de fusión Cas9 puede incluir un represor transcripcional, tal como el represor KRAB. La proteína de fusión Cas9 puede ser dCas9-KRAB. El dCas9-KRAB también puede afectar a la regulación de genes epigenéticos mediante el reclutamiento de factores formadores de heterocromatina en el locus diana. El sistema CRISPR/dCas9-KRAB puede usarse para reprimir la transcripción de genes, pero también puede usarse para dirigirse a elementos reguladores genómicos que antes eran inaccesibles mediante métodos de represión tradicionales como la interferencia de ARN (Fig. 53B). El suministro de dCas9-KRAB con ARNg dirigidos a un potenciador distal puede alterar la expresión de múltiples genes regulados por el potenciador diana (véase la Fig. 53C). El potenciador diana puede ser cualquier potenciador de un gen tal como el potenciador HS2.

45 a. Uso en métodos de transdiferenciación o diferenciación inducida

La presente divulgación proporciona un mecanismo para transfundir o inducir la diferenciación de células activando genes endógenos a través de ARN usando un sistema basado en CRISPR/Cas9, como se describió anteriormente.

(1) Transdiferenciación

55 El sistema basado en CRISPR/Cas9 puede usarse para transdiferenciar células. La transdiferenciación, también conocida como reprogramación de linaje o conversión directa, es un proceso donde las células se convierten de un tipo de célula diferenciada a otro sin pasar por un estado pluripotente intermedio o tipo de célula progenitora. Es un tipo de metaplasia, que incluye todos los interruptores de destino celular, incluyendo la interconversión de células madre. La transdiferenciación de células tiene usos potenciales para el modelado de enfermedades, el descubrimiento de fármacos, la terapia génica y la medicina regenerativa. La activación de genes endógenos, tales como *BRN2*, *MYT1L*, *ASCL1*, *NANOG* y/o *MYOD1*, usando el sistema basado en CRISPR/Cas9 descrito anteriormente puede
60 conducir a la transdiferenciación de varios tipos celulares, tales como fibroblastos, cardiomiositos, hepatocitos, condrocitos, células progenitoras mesenquimales, células madre hematopoyéticas o células de músculo liso, en fenotipos neuronales y miogénicos, respectivamente.

(2) Inducir la diferenciación

65 El sistema basado en CRISPR/Cas9 puede usarse para inducir la diferenciación de células, tales como células madre,

cardiomiositos, hepatocitos, condrocitos, células progenitoras mesenquimales, células madre hematopoyéticas o células de músculo liso. Por ejemplo, las células madre, tales como las células madre embrionarias o las células madre pluripotentes, pueden inducirse para diferenciarse en células musculares o células endoteliales vasculares, es decir, inducir la diferenciación neuronal o miogénica.

- 5 **12. Usos del sistema multiplex basado en CRISPR/Cas9**
- El sistema multiplex basado en CRISPR/Cas9 aprovecha la simplicidad y el bajo coste del diseño de ARN_{gm} y puede ser útil para explotar los avances en la investigación genómica de alto rendimiento usando la tecnología CRISPR/Cas9.
- 10 Por ejemplo, los vectores lentivíricos individuales descritos aquí son útiles para expresar Cas9 y numerosos ARN_{gm} en líneas celulares difíciles, tales como los fibroblastos primarios descritos aquí (Fig. 47). El sistema multiplex basado en CRISPR/Cas9 puede usarse de la misma manera que el sistema basado en CRISPR/Cas9 descrito anteriormente.

- 15 Además de la activación transcripcional y la funcionalidad de nucleasa descritas, este sistema será útil para expresar otros efectores novedosos basados en Cas9 que controlan modificaciones epigenéticas para diversos fines, incluyendo el interrogatorio de la arquitectura del genoma y las vías de regulación genética endógena. Como la regulación genética endógena es un delicado equilibrio entre múltiples enzimas, la multiplexación de sistemas Cas9 con diferentes funcionalidades permitirá examinar la compleja interacción entre diferentes señales regulatorias. El vector descrito en el presente documento debe ser compatible con ARN_{gm} modificados con aptámero y Cas9 ortogonales para permitir manipulaciones genéticas independientes usando un único conjunto de ARN_{gm}.

- 20 El sistema múltiple basado en CRISPR/Cas9 puede usarse para activar al menos un gen endógeno en una célula. El método incluye poner en contacto una célula con el vector lentivírico modificado. El gen endógeno puede activarse transitoriamente o activarse de forma estable. El gen endógeno puede estar reprimido transitoriamente o reprimido de manera estable. La proteína de fusión puede expresarse a niveles similares a los ARN_{gm}. La proteína de fusión se puede expresar a niveles diferentes en comparación con los ARN_{gm}. La célula puede ser una célula humana primaria.

- 25 El sistema multiplex basado en CRISPR/Cas9 puede usarse en un método de edición génica múltiple en una célula. El método incluye poner en contacto una célula con el vector lentivírico modificado. La edición génica múltiple puede incluir corregir un gen mutante o insertar un transgén. Corregir un gen mutante puede incluir eliminar, reorganizar o reemplazar el gen mutante. La corrección del gen mutante puede incluir la unión de extremos no homólogos mediada por nucleasa o la reparación dirigida por homología. La edición génica múltiple puede incluir eliminar o corregir al menos un gen, en donde el gen es un gen endógeno normal o un gen mutante. La edición génica múltiple puede incluir eliminar o corregir al menos dos genes. Por ejemplo, al menos dos genes, al menos tres genes, al menos cuatro genes, al menos cinco genes, al menos seis genes, al menos siete genes, al menos ocho genes, al menos nueve genes o al menos diez genes pueden eliminarse o corregirse.

- 30 El sistema basado en CRISPR/Cas9 múltiple puede usarse en un método de modulación múltiple de la expresión génica en una célula. El método incluye poner en contacto una célula con el vector lentivírico modificado. El método puede incluir modular los niveles de expresión génica de al menos un gen. La expresión génica de al menos un gen diana se modula cuando los niveles de expresión génica del al menos un gen diana aumentan o disminuyen en comparación con los niveles de expresión génica normales para el al menos un gen diana. Los niveles de expresión génica pueden ser niveles de ARN o proteínas.

45 **13. Uso en métodos para corregir un gen mutante y tratar a un sujeto**

- 50 La presente divulgación también es para su uso en un método para corregir un gen mutante en un sujeto. El uso comprende administrar al músculo esquelético o al músculo cardíaco del sujeto la composición para la edición genómica en el músculo esquelético o al músculo cardíaco, como se describió anteriormente. El uso de la composición para suministrar la nucleasa específica del sitio al músculo esquelético o al músculo cardíaco puede restaurar la expresión de una proteína completamente funcional o parcialmente funcional con una plantilla de reparación o ADN donante, que puede reemplazar el gen completo o la región que contiene la mutación. Puede usarse la nucleasa específica de sitio para introducir roturas bícatenarias específicas de sitio en loci genómicos marcados como diana. Las roturas bícatenarias específicas del sitio se crean cuando la nucleasa específica del sitio se une a una secuencia de ADN diana, permitiendo de esta manera la escisión del ADN diana. Esta escisión del ADN puede estimular la maquinaria natural de reparación del ADN, conduciendo a una de dos posibles vías de reparación: reparación dirigida por homología (HDR) o la vía de unión de extremos no homólogos (NHEJ).

- 55 La presente divulgación se dirige a la edición genómica con una nucleasa específica de sitio sin una plantilla de reparación, que puede corregir eficientemente el marco de lectura y restaurar la expresión de una proteína funcional implicada en una enfermedad genética. Las nucleasas específicas de sitio divulgadas pueden implicar el uso de enfoques de corrección basados en reparación dirigida por homología o unión de extremos no homólogos mediada por nucleasas (NHEJ), que permiten una corrección eficiente en líneas celulares primarias con proliferación limitada que pueden no ser susceptibles de recombinación homóloga o corrección genética basada en selección. Esta estrategia integra el ensamblaje rápido y robusto de nucleasas activas de sitio específico con un método eficiente de edición génica para el tratamiento de enfermedades genéticas provocadas por mutaciones en regiones codificantes

no esenciales que provocan desplazamientos de marco, codones de parada prematuros, sitios donantes de corte y empalme aberrantes o sitios aceptores de corte y empalme aberrantes.

a. Unión de extremos no homólogos mediada por nucleasa

La restauración de la expresión de proteínas a partir de un gen mutado endógeno puede realizarse mediante la reparación del ADN mediada por NHEJ sin plantilla. A diferencia de un método transitorio dirigido al ARN del gen diana, la corrección del marco de lectura del gen diana en el genoma mediante una nucleasa específica de sitio expresada transitoriamente puede conducir a una expresión del gen diana restaurada permanentemente por cada célula modificada y toda su progenie.

La corrección del gen NHEJ mediada por nucleasa puede corregir el gen diana mutado y ofrece varias ventajas potenciales sobre la vía HDR. Por ejemplo, NHEJ no requiere una plantilla de donante, que puede provocar mutagénesis de inserción inespecífica. A diferencia de HDR, NHEJ funciona eficientemente en todas las etapas del ciclo celular y, por lo tanto, puede explotarse eficazmente tanto en células cíclicas como postmitóticas, tales como las fibras musculares. Esto proporciona una robusta, alternativa de restauración genética permanente a la omisión de exones basada en oligonucleótidos o la lectura farmacológica forzada de codones de parada y, en teoría, podría requerir tan solo un tratamiento farmacológico. La corrección genética basada en NHEJ usando un sistema basado en CRISPR/Cas9, así como otras nucleasas diseñadas incluyendo meganucleasas y nucleasas con dedos de cinc, puede combinarse con otras plataformas *ex vivo* e *in vivo* existentes para terapias basadas en células y genes, además del enfoque de electroporación de plásmidos descrito aquí. Por ejemplo, el suministro de un sistema basado en CRISPR/Cas9 mediante transferencia de genes basada en ARNm o como proteínas permeables a células purificadas podría permitir un enfoque de edición genómica libre de ADN que evitaría cualquier posibilidad de mutagénesis por inserción.

b. Reparación dirigida por homología

La restauración de la expresión de proteínas a partir de un gen mutado endógeno puede implicar una reparación dirigida por homología. El método descrito anteriormente incluye además la administración de una plantilla de donante a la célula. La plantilla donante puede incluir una secuencia nucleotídica que codifica una proteína completamente funcional o una proteína parcialmente funcional. Por ejemplo, la plantilla donante puede incluir una construcción de distrofina miniaturizada, denominada minidistrofina ("minidys"), una construcción de distrofina completamente funcional para restaurar un gen de distrofina mutante, o un fragmento del gen de distrofina que después de la reparación dirigida por homología conduce a la restauración del gen de distrofina mutante.

c. Uso en métodos para corregir un gen mutante y tratar a un sujeto usando CRISPR/Cas9

La presente divulgación también se dirige a la edición genómica con el sistema basado en CRISPR/Cas9 para restaurar la expresión de una proteína completamente funcional o parcialmente funcional con una plantilla de reparación o ADN donante, que puede reemplazar el gen completo o la región que contiene la mutación. El sistema basado en CRISPR/Cas9 puede usarse para introducir roturas biconicinas específicas de sitio en loci genómicos marcados como diana. Las roturas biconicinas específicas del sitio se crean cuando el sistema basado en CRISPR/Cas9 se une a una secuencia de ADN diana usando el ARNg, permitiendo de esta manera la escisión del ADN diana. El sistema basado en CRISPR/Cas9 tiene la ventaja de la edición avanzada del genoma debido a su alta tasa de modificación génica exitosa y eficiente. Esta escisión del ADN puede estimular la maquinaria natural de reparación del ADN, conduciendo a una de dos posibles vías de reparación: reparación dirigida por homología (HDR) o la vía de unión de extremos no homólogos (NHEJ). Por ejemplo, un sistema basado en CRISPR/Cas9 dirigido al gen de la distrofina puede incluir un ARNg que tiene una secuencia de ácido nucleico de cualquiera de SEQ ID NO: 65-115.

La presente divulgación se dirige a la edición genómica con un sistema basado en CRISPR/Cas9 sin una plantilla de reparación, que puede corregir eficientemente el marco de lectura y restaurar la expresión de una proteína funcional implicada en una enfermedad genética. El sistema y los métodos basados en CRISPR/Cas9 divulgados pueden implicar el uso de reparación dirigida por homología o enfoques de corrección basados en unión de extremos no homólogos mediada por nucleasa (NHEJ), que permiten una corrección eficiente en líneas celulares primarias con proliferación limitada que pueden no ser susceptibles de recombinación homóloga o corrección genética basada en selección. Esta estrategia integra el ensamblaje rápido y robusto de un sistema activo basado en CRISPR/Cas9 con un método eficiente de edición génica para el tratamiento de enfermedades genéticas provocadas por mutaciones en regiones codificantes no esenciales que provocan desplazamientos de marco, codones de parada prematuros, sitios donantes de corte y empalme aberrantes o sitios aceptores de corte y empalme aberrantes.

La presente divulgación es para su uso en métodos para corregir un gen mutante en una célula y tratar a un sujeto que padece una enfermedad genética, tal como DMD. El uso puede incluir la administración a una célula o sujeto de un sistema basado en CRISPR/Cas9, un polinucleótido o vector que codifica dicho sistema basado en CRISPR/Cas9 o una composición de dicho sistema basado en CRISPR/Cas9 como se describió anteriormente. El uso puede incluir la administración de un sistema basado en CRISPR/Cas9, tal como administrar una proteína Cas9 o una proteína de

fusión Cas9 que contiene un segundo dominio que tiene actividad nucleasa, una secuencia nucleotídica que codifica dicha proteína Cas9 o proteína de fusión Cas9 y/o al menos un ARNg, en donde los ARNg se dirigen a diferentes secuencias de ADN. Las secuencias de ADN diana pueden estar superpuestas. El número de ARNg administrados a la célula puede ser al menos 1 ARNg, al menos 2 ARNg diferentes, al menos 3 ARNg diferentes, al menos 4 ARNg diferentes, al menos 5 ARNg diferentes, al menos 6 ARNg diferentes, al menos 7 ARNg diferentes, al menos 8 ARNg diferentes, al menos 9 ARNg diferentes, al menos 10 ARNg diferentes, al menos 15 ARNg diferentes, al menos 20 ARNg diferentes, al menos 30 ARNg diferentes o al menos 50 ARNg diferentes, como se describió anteriormente. El ARNg puede incluir una secuencia de ácido nucleico de al menos una de SEQ ID NO: 65-115. El método puede implicar reparación dirigida por homología o unión de extremos no homólogos.

10 14. Uso en métodos para tratar una enfermedad

La presente divulgación es para su uso en un método para tratar a un sujeto que lo necesita. El uso comprende administrar a un tejido de un sujeto la composición para alterar la expresión génica y diseñar o alterar el ADN genómico en una célula o editar el genoma del sujeto, como se describió anteriormente. El uso puede comprender administrar al músculo esquelético o al músculo cardíaco del sujeto la composición para la edición genómica en el músculo esquelético o al músculo cardíaco, como se describió anteriormente. El sujeto puede sufrir una afección del músculo esquelético o del músculo cardíaco que provoca degeneración o debilidad o una enfermedad genética. Por ejemplo, el sujeto puede estar padeciendo distrofia muscular de Duchenne, como se describió anteriormente.

20 a. Distrofia muscular de Duchenne

La divulgación como se describe anteriormente, puede usarse para corregir el gen de la distrofina y recuperar la expresión proteica totalmente funcional o parcialmente funcional de dicho gen de distrofina mutado. En algunos aspectos la divulgación es para su uso en la reducción de los efectos (por ejemplo, síntomas/indicaciones clínicas) de DMD en un paciente. En algunos aspectos la divulgación es para su uso en el tratamiento de DMD en un paciente. En algunos aspectos la divulgación es para su uso en la prevención de DMD en un paciente. En algunos aspectos la divulgación es para su uso en la prevención de avance adicional de DMD en un paciente.

30 15. Construcciones y plásmidos

Las composiciones, como se ha descrito anteriormente, pueden comprender construcciones genéticas que codifican el sistema basado en CRISPR/Cas9, como se divulga en el presente documento. La construcción genética, tal como un plásmido, puede comprender un ácido nucleico que codifica el sistema basado en CRISPR/Cas9, tales como la proteína Cas9 y las proteínas de fusión Cas9 y/o al menos uno de los ARNg. Las composiciones, como se ha descrito anteriormente, pueden comprender construcciones genéticas que codifican el vector AAV modificado y una secuencia de ácido nucleico que codifica la nucleasa específica del sitio, como se divulga en el presente documento. La construcción genética, tal como un plásmido, puede comprender un ácido nucleico que codifica la nucleasa específica de sitio. Las composiciones, como se ha descrito anteriormente, pueden comprender construcciones genéticas que codifican el vector lentivírico modificado, como se divulga en el presente documento. La construcción genética, tal como un plásmido, puede comprender un ácido nucleico que codifica la proteína de fusión Cas9 y al menos un ARNg. La construcción genética puede estar presente en la célula como una molécula extracromosómica funcional. La construcción genética puede ser un minicromosoma lineal que incluye centrómero, telómeros o plásmidos o cósmidos.

45 La construcción genética también puede ser parte de un genoma de un vector vírico recombinante, incluyendo lentivirus recombinantes, adenovirus recombinantes y virus asociados a adenovirus recombinantes. La construcción genética puede ser parte del material genético de microorganismos vivos atenuados o vectores microbianos recombinantes que viven en células. Las construcciones genéticas pueden comprender elementos reguladores para la expresión génica de las secuencias codificantes del ácido nucleico. Los elementos reguladores pueden ser un promotor, un potenciador, un codón de iniciación, un codón de parada o una señal de poliadenilación.

50 Las secuencias de ácido nucleico pueden formar una construcción genética que puede ser un vector. El vector puede ser capaz de expresar la proteína de fusión, tal como la proteína de fusión Cas9 o la nucleasa específica de sitio, en la célula de un mamífero. El vector puede ser recombinante. El vector puede comprender ácido nucleico heterólogo que codifica la proteína de fusión, tal como la proteína de fusión Cas9 o la nucleasa específica de sitio. El vector puede ser un plásmido. El vector puede ser útil para transfectar células con ácido nucleico que codifica la proteína de fusión Cas9 o nucleasa específica de sitio, en donde la célula huésped transformada se cultiva y se mantiene en condiciones en las que tiene lugar la expresión de la proteína de fusión Cas9 o del sistema de nucleasa específico de sitio.

60 Las secuencias codificantes pueden optimizarse para lograr estabilidad y altos niveles de expresión. En algunos casos, los codones se seleccionan para reducir la formación de estructuras secundarias del ARN tales como la que se forma debido al enlace intramolecular.

65 El vector puede comprender ácido nucleico heterólogo que codifica el sistema basado en CRISPR/Cas9 o la nucleasa específica de sitio y puede comprender además un codón de iniciación, que puede estar en dirección 5' del sistema basado en CRISPR/Cas9 o la secuencia codificante de nucleasa específica del sitio, y un codón de parada, que puede

- estar en dirección 3' del sistema basado en CRISPR/Cas9 o de la secuencia codificante de nucleasa específica del sitio. El codón de iniciación y terminación puede estar en marco con el sistema basado en CRISPR/Cas9 o la secuencia codificante de nucleasa específica del sitio. El vector también puede comprender un promotor que está operativamente unido al sistema basado en CRISPR/Cas9 o a la secuencia codificante de nucleasa específica del sitio. El promotor unido operativamente al sistema basado en CRISPR/Cas9 o la secuencia codificante de nucleasa específica del sitio puede ser un promotor del virus 40 del simio (SV40), un promotor del virus del tumor mamario de ratón (MMTV), un promotor del virus de la inmunodeficiencia humana (VIH) tal como el promotor de repetición terminal larga (LTR) del virus de la inmunodeficiencia bovina (BIV), un promotor del virus Moloney, un promotor del virus de la leucosis aviar (ALV), un promotor de citomegalovirus (CMV) tal como el promotor temprano inmediato de CMV, un promotor del virus de Epstein Barr (EBV) o un promotor del virus del sarcoma de Rous (RSV). El promotor también puede ser un promotor de un gen humano tal como ubiquitina C humana (hUbC), actina humana, miosina humana, hemoglobina humana, creatina muscular humana o metalotioneína humana. El promotor también puede ser un promotor específico de tejido, tal como un promotor específico de músculo o piel, natural o sintético. Algunos ejemplos de dichos promotores se describen en la Publicación de Solicitud de Patente de EE.UU. N.º US20040175727.
- El vector también puede comprender una señal de poliadenilación, que puede estar en dirección 3' del sistema basado en CRISPR/Cas9 o de la nucleasa específica del sitio. La señal de poliadenilación puede ser una señal de poliadenilación de SV40, una señal de poliadenilación LTR, una señal de poliadenilación de la hormona del crecimiento bovino (bGH), una señal de poliadenilación de la hormona de crecimiento humana (hGH) o una señal de poliadenilación de β-globina humana. La señal de poliadenilación de SV40 puede ser una señal de poliadenilación de un vector pCEP4 (Invitrogen, San Diego, CA).
- El vector también puede comprender un potenciador en dirección 5' del sistema basado en CRISPR/Cas9, es decir, la secuencia codificante de la proteína Cas9 o de la proteína de fusión Cas9 o ARNgm, o la nucleasa específica del sitio. El potenciador puede ser necesario para la expresión del ADN. El potenciador puede ser actina humana, miosina humana, hemoglobina humana, creatina muscular humana o un potenciador vírico tal como uno de CMV, HA, RSV o EBV. Los potenciadores de la función de los polinucleótidos se describen en las Patentes de EE.UU. N.º 5.593.972, 5.962.428 y el documento WO94/016737. El vector también puede comprender un origen de replicación de mamífero para mantener el vector extracromosómicamente y producir múltiples copias del vector en una célula. El vector también puede comprender una secuencia reguladora, que puede ser muy adecuada para la expresión génica en una célula de mamífero o humana en la que se administra el vector. El vector también puede comprender un gen indicador, tal como proteína fluorescente verde ("GFP") y/o un marcador seleccionable, tal como higromicina ("Hygro").
- El vector puede ser vectores de expresión o sistemas para producir proteínas mediante técnicas rutinarias y materiales de partida fácilmente disponibles incluyendo Sambrook *et al.*, Molecular Cloning and Laboratory Manual, Segunda Ed., Cold Spring Harbor (1989). El vector puede comprender la secuencia de ácido nucleico que codifica el sistema basado en CRISPR/Cas9, incluyendo la secuencia de ácido nucleico que codifica la proteína Cas9 o la proteína de fusión Cas9 y la secuencia de ácido nucleico que codifica al menos un ARNg que comprende la secuencia de ácido nucleico de al menos una de SEQ ID NO: 5-40, 65-144, 492-515, 540-563 y 585-625.
- ## 16. Composiciones farmacéuticas
- La composición puede estar en una composición farmacéutica. La composición farmacéutica puede comprender de aproximadamente 1 ng a aproximadamente 10 mg de ADN que codifica el sistema basado en CRISPR/Cas9 o el componente proteico del sistema basado en CRISPR/Cas9, es decir, la proteína Cas9 o proteína de fusión Cas9. La composición farmacéutica puede comprender de aproximadamente 1 ng a aproximadamente 10 mg del ADN del vector AAV modificado y la secuencia nucleotídica que codifica la nucleasa específica de sitio. La composición farmacéutica puede comprender de aproximadamente 1 ng a aproximadamente 10 mg del ADN del vector lentivírico modificado. Las composiciones farmacéuticas se formulan de acuerdo con el modo de administración a usar. En los casos en que las composiciones farmacéuticas sean composiciones farmacéuticas inyectables, son estériles, apirógenas y sin partículas. Preferentemente se usa una formulación isotónica. Generalmente, los aditivos para la isotonicidad pueden incluir cloruro sódico, dextrosa, manitol, sorbitol y lactosa. En algunos casos, se prefieren soluciones isotónicas tales como solución salina tamponada con fosfato. Los estabilizadores incluyen gelatina y albúmina. Puede añadirse un agente vasoconstrictor a la formulación.
- La composición puede comprender además un excipiente farmacéuticamente aceptable. El excipiente farmacéuticamente aceptable puede ser moléculas funcionales como vehículos, adyuvantes, portadores o diluyentes. El excipiente farmacéuticamente aceptable puede ser un agente facilitador de la transfección, que puede incluir agentes tensioactivos, tales como los complejos inmunoestimulantes (SCOMS), adyuvante incompleto de Freund, análogo de LPS que incluye monofosforil lípido A, péptidos de muramilo, análogos de quinona, vesículas tales como escualano y escualeno, ácido hialurónico, lípidos, liposomas, iones calcio, proteínas víricas, polianiones, policationes o nanopartículas u otros agentes facilitadores de la transfección conocidos.
- El agente facilitador de la transfección es un polianión, un polication, incluyendo poli-L-glutamato (LGS) o un lípido. El agente facilitador de la transfección es poli-L-glutamato y, más preferentemente, el poli-L-glutamato está presente en la composición para la edición genómica en músculo esquelético o músculo cardíaco en una concentración de menos

de 6 mg/ml. El agente facilitador de la transfección también puede incluir agentes tensioactivos tales como complejos inmunoestimulantes (ISCOMS), adyuvante incompleto de Freund, análogo de LPS que incluye monofosforil lípido A, péptidos de muramilo, análogos de quinona y vesículas tales como escualano y escualeno y ácido hialurónico también pueden usarse administrados junto con la construcción genética. El vector de ADN que codifica la composición también 5 puede incluir un agente facilitador de la transfección tales como lípidos, liposomas, incluyendo liposomas de lecitina u otros liposomas conocidos en la técnica, como una mezcla de ADN-liposoma (véase, por ejemplo, el documento WO9324640), iones calcio, proteínas víricas, polianiones, policationes nanopartículas u otros agentes facilitadores 10 de la transfección conocidos. Preferentemente, el agente facilitador de la transfección es un polanión, un policación, incluyendo poli-L-glutamato (LGS) o un lípido.

17. Métodos de suministro

En el presente documento se proporciona un método para suministrar las formulaciones farmacéuticas, preferentemente las composiciones descritas anteriormente, para proporcionar construcciones genéticas. El 15 suministro de las composiciones puede ser la transfección o la electroporación de la composición como una molécula de ácido nucleico que se expresa en la célula y se suministra a la superficie de la célula. Las moléculas de ácido nucleico pueden electroporarse usando dispositivos BioRad Gene Pulser Xcell o Amaxa Nucleofector lib. Pueden usarse varios tampones diferentes, incluyendo solución de electroporación BioRad, solución salina tamponada con fosfato de Sigma Producto n.º D8537 (PBS), Invitrogen OptiMEM I (OM) o Amaxa Nucleofector solution V (N. V.). Las 20 transfecciones pueden incluir un reactivo de transfección, tales como Lipofectamine 2000.

Tras el suministro de la composición al tejido y posteriormente el vector a las células del mamífero, las células 25 transfectadas expresarán la proteína de fusión, tal como un sistema basado en CRISPR/Cas9 y/o una nucleasa específica de sitio. La composición puede administrarse a un mamífero para alterar la expresión génica o para rediseñar o alterar el genoma. Por ejemplo, la composición puede administrarse a un mamífero para corregir el gen de la distrofina en un mamífero. El mamífero puede ser humano, primate no humano, vaca, cerdo, oveja, cabra, antílope, bisonte, búfalo de agua, bóvidos, ciervo, erizos, elefantes, llama, alpaca, ratones, ratas o pollos y, preferentemente humano, vaca, cerdo o pollo.

a. Sistema basado en CRISPR/Cas9

El vector que codifica un componente proteico del sistema basado en CRISPR/Cas9, es decir, la proteína Cas9 o proteína de fusión Cas9, puede suministrarse al mamífero mediante inyección de ADN (también denominada vacunación de ADN) con y sin electroporación *in vivo*, mediada por liposomas, facilitada por nanopartículas y/o 30 vectores recombinantes. El vector recombinante puede suministrarse mediante cualquier modo vírico. El modo vírico puede ser un lentivirus recombinante, un adenovirus recombinante y/o un virus adenoasociado recombinante.

El nucleótido que codifica un componente proteico del sistema basado en CRISPR/Cas9, es decir, la proteína Cas9 o 40 proteína de fusión Cas9, puede introducirse en una célula para corregir genéticamente el gen diana o alterar la expresión génica de un gen, tal como activar o reprimir genes endógenos. Por ejemplo, un nucleótido que codifica un componente proteico del sistema basado en CRISPR/Cas9, es decir, la proteína Cas9 o proteína de fusión Cas9, dirigido hacia un gen de distrofina mutante por el ARNg puede introducirse en una célula de mioblasto de un paciente con DMD. Como alternativa, pueden introducirse en una célula de fibroblastos de un paciente con DMD, y la célula de 45 fibroblastos corregida genéticamente puede tratarse con MyoD para inducir la diferenciación en mioblastos, que pueden implantarse en sujetos, tales como los músculos dañados de un sujeto para verificar que la proteína distrofina corregida era funcional y/o para tratar al sujeto. Las células modificadas también pueden ser células madre, tales como las células madre pluripotentes inducidas, progenitores derivados de la médula ósea, progenitores del músculo esquelético, mioblastos esqueléticos humanos de pacientes con DMD, células CD133+, mesoangioblastos y células transducidas con MyoD o Pax7 u otras células progenitoras miogénicas. Por ejemplo, el sistema basado en 50 CRISPR/Cas9 puede provocar la diferenciación neuronal o miogénica de una célula madre pluripotente inducida.

18. Rutas de administración

Las composiciones pueden administrarse a un sujeto por diferentes vías incluyendo por vía oral, por vía parenteral, 55 por vía sublingual, por vía transdérmica, por vía rectal, por vía transmucosa, por vía tópica, por inhalación, por administración bucal, por vía intrapleural, por vía intravenosa, por vía intraarterial, por vía intraperitoneal, por vía subcutánea, por vía intramuscular, por vía intranasal, por vía intratecal y por vía intraarticular o combinaciones de los mismos. Para uso veterinario, la composición puede administrarse como una formulación adecuadamente aceptable de acuerdo con la práctica veterinaria normal. El veterinario puede determinar fácilmente el régimen de dosificación y 60 la vía de administración más apropiada para un animal en particular. Las composiciones pueden administrarse mediante jeringas tradicionales, dispositivos de inyección sin aguja, "pistolas de bombardeo de microproyectiles" u otros métodos físicos tales como electroporación ("EP"), "método hidrodinámico" o ultrasonidos.

La composición puede suministrarse al mamífero mediante diversas tecnologías que incluyen inyección de ADN 65 (también denominada vacunación con ADN) con y sin electroporación *in vivo*, mediada por liposomas, facilitada por nanopartículas, vectores recombinantes tales como lentivirus recombinante, adenovirus recombinantes y virus

asociados a adenovirus recombinantes. La composición puede inyectarse en el músculo esquelético o el músculo cardíaco. Por ejemplo, la composición puede inyectarse en el músculo tibial anterior.

19. Tipos de células

5 Cualquiera de estos métodos de suministro y/o vías de administración podría utilizarse con una infinidad de tipos de células, por ejemplo, aquellos tipos de células que actualmente se están investigando para terapias basadas en células. Los tipos celulares pueden ser fibroblastos, células madre pluripotentes, cardiomiositos, hepatocitos, condrocitos, células progenitoras mesenquimales, células madre hematopoyéticas, células de músculo liso o línea 10 celular de leucemia eritroide humana K562.

a. DMD

15 Los tipos de células que actualmente se están investigando para terapias celulares de DMD incluyen células de mioblastos inmortalizados, tales como líneas derivadas de pacientes con DMD y de tipo silvestre, por ejemplo las líneas celulares Δ48-50 DMD, DMD 8036 (del48-50), C25C14 y DMD-7796, fibroblastos dérmicos DMD primarios, células madre pluripotentes inducidas, progenitores derivados de la médula ósea, progenitores del músculo esquelético, mioblastos esqueléticos humanos de pacientes con DMD, células CD 133+, mesoangioblastos, cardiomiositos, hepatocitos, condrocitos, células progenitoras mesenquimales, células madre hematopoyéticas, células de músculo liso y células transducidas con MyoD o Pax7 u otras células progenitoras miogénicas. La inmortalización de células miogénicas humanas puede usarse para la derivación clonal de células miogénicas genéticamente corregidas. Las células pueden modificarse *ex-vivo* para aislar y expandir poblaciones clonales de mioblastos DMD inmortalizados que contienen un gen de distrofina genéticamente corregido y están libres de otras mutaciones introducidas por nucleasas en regiones codificantes de proteínas del genoma. Como alternativa, el suministro *in vivo* transitorio de nucleasas mediante transferencia de genes no víricos o no integrantes víricos o mediante suministro directo de proteínas purificadas y ARN que contienen motivos de penetración celular, puede permitir una corrección altamente específica *in situ* con riesgo mínimo o nulo de integración del ADN exógeno.

20

25

20. Kits

30 Se divulga en el presente documento un kit, que puede usarse para editar un genoma en el músculo esquelético o en el músculo cardíaco, tal como corregir un gen mutante. El kit comprende una composición para la edición genómica en músculo esquelético o músculo cardíaco, como se describió anteriormente, e instrucciones para usar dicha composición. Las instrucciones incluidas en los kits pueden estar adheridas al material de envasado o pueden incluirse como un prospecto. Aunque las instrucciones normalmente son materiales escritos o impresos no se limitan a estos. La presente divulgación contempla cualquier medio capaz de almacenar dichas instrucciones y de comunicárselas a un usuario final. Dichos medios incluyen, pero no se limitan a, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, CD ROM) y similares. Como se usa en el presente documento, el término "instrucciones" puede incluir la dirección de un sitio de Internet que proporciona las 40 instrucciones.

45 La composición para la edición genómica en músculo esquelético o el músculo cardíaco puede incluir un vector AAV modificado y una secuencia nucleotídica que codifica una nucleasa específica de sitio, como se describió anteriormente. La nucleasa específica de sitio puede incluir una ZFN, una TALEN o un sistema basado en CRISPR/Cas9, como se ha descrito anteriormente, que se une a y escinde específicamente un gen mutado. La nucleasa específica de sitio, como se ha descrito anteriormente, puede incluirse en el kit para unirse específicamente y dirigirse a una región particular en el gen mutado. La nucleasa específica de sitio puede ser específica para un gen de distrofina mutado, como se describió anteriormente. El kit puede incluir además ADN donante, un ARN que 50 transgén, como se describió anteriormente.

a. Sistema basado en CRISPR/Cas9

55 Se divulga en el presente documento un kit, que puede usarse para corregir un gen mutado. El kit comprende al menos un componente para corregir un gen mutado e instrucciones para usar el sistema basado en CRISPR/Cas9. Las instrucciones incluidas en los kits pueden estar adheridas al material de envasado o pueden incluirse como un prospecto. Aunque las instrucciones normalmente son materiales escritos o impresos no se limitan a estos. La presente divulgación contempla cualquier medio capaz de almacenar dichas instrucciones y de comunicárselas a un usuario final. Dichos medios incluyen, pero no se limitan a, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, CD ROM) y similares. Como se usa en el presente 60 documento, el término "instrucciones" puede incluir la dirección de un sitio de Internet que proporciona las instrucciones.

65 Al menos un componente puede incluir al menos un sistema basado en CRISPR/Cas9, como se ha descrito anteriormente, que se dirige específicamente a un gen. El kit puede incluir una proteína Cas9 o una proteína de fusión Cas9, una secuencia nucleotídica que codifica dicha proteína Cas9 o proteína de fusión Cas9 y/o al menos un ARN. El sistema basado en CRISPR/Cas9, como se ha descrito anteriormente, puede incluirse en el kit para unirse

específicamente y dirigirse a una región diana particular en dirección 5', dentro de o en dirección 3' de la región codificante del gen diana. Por ejemplo, un sistema basado en CRISPR/Cas9 puede ser específico para una región promotor de un gen diana o un sistema basado en CRISPR/Cas9 puede ser específico para un gen mutado, por ejemplo, un gen de distrofina mutado, como se describió anteriormente. El kit puede incluir ADN de donante, como se describió anteriormente.

21. Ejemplos

Habiendo descrito ahora la presente divulgación en detalle, lo mismo se entenderá más claramente con referencia a los siguientes ejemplos, que meramente pretenden ilustrar algunos aspectos y realizaciones de la divulgación, y no deben considerarse limitantes del alcance de la divulgación.

La presente invención tiene múltiples aspectos, ilustrados por los siguientes ejemplos no limitantes.

15 Ejemplo 1

Materiales y Métodos

Cultivo celular y transfección. Las células HEK293T se obtuvieron del American Tissue Collection Center (ATCC) a través de las instalaciones del Centro Oncológico de la Universidad de Duke y se mantuvieron en DMEM suplementado con suero bovino fetal al 10 % y penicilina/estreptomicina al 1 % a 37 °C con CO₂ al 5 %. Las células HEK293T se transfecaron con Lipofectamine 2000 (Invitrogen) de acuerdo con las instrucciones del fabricante. Las eficiencias de transfección fueron habitualmente superiores al 80 % como se determina mediante microscopía de fluorescencia tras el suministro de un plásmido de expresión de eGFP de control. El plásmido de expresión Cas9 se transfeció en una proporción de masa de 3:1 con los plásmidos de expresión de ARNg individuales o con la cantidad idéntica de plásmido de expresión de ARNg que consiste en una mezcla de cantidades iguales de los cuatro ARNg.

Los fibroblastos embrionarios primarios de ratón (PMEF-HL, Millipore, Billerica, MA) se sembraron (75.000 por pocillo) en placas TCPS de 24 pocillos (BD, Franklin Lakes, NJ) y se mantuvieron a 37 °C y CO₂ al 5 % en medio MEF completo que consiste en DMEM con alto contenido de glucosa suplementado con Premium Select FBS al 10 % (Atlanta Biologicals, Lawrenceville, GA), gentamicina 25 µg ml⁻¹ (Invitrogen), 1X GlutaMAX, aminoácidos no esenciales, piruvato sódico y β-mercaptopetanol (Invitrogen). Las transfecciones MEF se realizaron con una dosis individual de 1 µg cm⁻² de ADN plasmídico total, suministrado como nanocomplejos catiónicos después de la condensación electrostática con polí(CBA-ABOL) en OptiMEM sin suero ni antibióticos, como se describió anteriormente (Adler *et al.* *Molecular therapy. Nucleic acids* 1, e32 (2012)). OptiMEM se reemplazó por medio MEF completo 4 h después de la transfección. 48 h después de la transfección, MEF se procesaron para qRT-PCR o el medio MEF completo se reemplazó por medio de inducción neural N3 que contenía: DMEM/F-12 (Invitrogen), Suplemento 1X N-2 (Invitrogen), bFGF2 humano 10 ng ml⁻¹ (Stemgent, Cambridge, MA) y gentamicina 25 µg ml⁻¹ (Invitrogen). Un vector indicador de GFP (pmax-GFP, 3486 pb, Amaxa, Colonia, Alemania) se usó para optimizar las condiciones de transfección. El plásmido de expresión Cas9 se transfeció en una proporción de masa de 3:1 o 1:1 a una mezcla igual de cuatro plásmidos de expresión de ARNg.

Plásmidos. Los plásmidos que codifican Cas9 de tipo silvestre y H840A se obtuvieron de Addgene (Plásmido N.º 39312 y Plásmido N.º 39316; Jinek, *et al.* *Science* 337, 816-821 (2012)). H840A Cas9 se clonó en el vector pcDNA3.1 en marco con una etiqueta de epítopo FLAG y una secuencia de localización nuclear (NLS) en el extremo N con un par de cebadores que introdujo la mutación D10A. El dominio VP64, un NLS y una etiqueta de epítopo HA se clonaron en marco con el ORF de Cas9 en el extremo C (Fig. 1a, Fig. 9a). Los casetes de expresión de ARNtracr y ARNcr (Cong *et al.* *Science* 339, 819-823 (2013)) se ordenaron como gBlocks (Integrated DNA Technologies (IDT)) y se clonaron en un plásmido pZDonor (Sigma) con sitios KpnI y SacII. Un casete de expresión de ARN guía químico (Mali *et al.* *Science* 339, 823-826 (2013)) también se ordenó como gBlocks con modificaciones para incluir un sitio de restricción BbsI para facilitar la clonación rápida de nuevas secuencias espaciadoras de ARN guía (Fig. 9b). Los oligonucleótidos que contienen las secuencias diana se obtuvieron de IDT, se hibridaron, se fosforilaron y se clonaron en los plásmidos apropiados usando sitios BbsI. Las secuencias diana se proporcionan en la Tabla 2.

55

Tabla 2 Secuencias diana de ARNg

Diana	Nombre	Secuencia	SEQ ID NO
ASCL1	CR1	GCTGGGTGCCATTGAAA	5
	CR2	CAGCCGCTCGCTGCAGCAG	6
	CR3	TGGAGAGTTGCAAGGAGC	7
	CR4	GTTCATTTCAGCCGGAGTC	8
NANOG	CR1	CGCCAGGAGGGTGGGTCTA	9
	CR2	CCTTGGTGAGACTGGTAGA	10
	CR3	GTCTTCAGGTTCTGTTGCT	11

	CR4	ATATTCTGATTTAAAAGT	12
VEGFA	CR1	TTAAAAGTCGGCTGGTAGC	13
	CR2	CGGGCCGGGGCGGGGTCC	14
	CR3	GCCCAGCCGCGTGTGGAA	15
	CR4	CCTTCATTGCAGCGGGCTG	16
TERT	CR1	CCGACCCCTCCGGGTCCC	17
	CR2	CAGGACCGCGCTTCCCACG	18
	CR3	TGCACCCTGGGAGCGCGAG	19
	CR4	CCGCACGCACCTGTTCCA	20
IL1B	CR1	AAAACAGCGAGGGAGAAC	21
	CR2	TTAACTTGATTGTGAAATC	22
	CR3	AAAACAATGCATATTGCA	23
	CR4	AAAATCCAGTATTTAATG	24
IL1R2	CR1	ACCCAGCACTGCAGCCTGG	25
	CR2	AACTTATGCAGCGTTCCCT	26
	CR3	TCACTTAAAACCACCTCT	27
	CR4	GCATCTTTCTCTTTAAT	28
IL1RN	CR1	TGTACTCTCTGAGGTGCTC	29
	CR2	ACGCAGATAAGAACCAAGTT	30
	CR3	CATCAAGTCAGCCATCAGC	31
	CR4	GAGTCACCCCTCCTGGAAAC	32
HBG1/2	CR1	GCTAGGGATGAAGAATAAA	33
	CR2	TTGACCAATAGCCTTGACA	34
	CR3	TGCAAATATCTGTCTGAAA	35
	CR4	AAATTAGCAGTATCCTCTT	36
MYOD1	CR1	CCTGGGCTCCGGGGCGTT	37
	CR2	GGCCCTGCGGCCACCCCG	38
	CR3	CTCCCTCCCTGCCCGTAG	39
	CR4	AGGTTGGAAAGGGCGTGC	40

Transferencia Western. Las células se lisaron en Tris-Cl 50 mM (pH 7,4), NaCl 150 mM, TritonX-100 al 0,5 % y SDS al 0,1 %. Los lisados se mezclaron con tampón de carga, se hirvieron durante 5 min y se procesaron volúmenes iguales de proteína en geles de poliacrilamida NuPAGE® Novex al 4-12 % o Bis-Tris Gel al 10 % y se transfirieron a membranas de nitrocelulosa. La unión del anticuerpo no específico se bloqueó con Tris 50 mM/NaCl 150 mM/Tween-

5 20 al 0,1 % (TBS-T) con leche desnatada al 5 % durante 30 min. Las membranas se incubaron con anticuerpos primarios (anti-Flag conjugado con HRP (Cell Signaling, N.º de Cat. 2044) en BSA al 5 % en TBS-T diluido 1:1000 durante la noche; anti-GAPDH (Cell Signaling, clon 1400) en leche al 5 % en TBS-T diluido 1:5000 durante 30 min; anti-ASCL1 (Santa Cruz, clon sc-48449) en BSA al 5 % diluido 1:500; o anti-g-globina (Santa Cruz, clon 51-7) en leche al 5 % diluida 1:500 y las membranas se lavaron con TBS-T durante 30 min. Las membranas marcadas con anticuerpos primarios se incubaron con anticuerpo anti-conejo conjugado con HRP (Sigma-Aldrich) diluido 1:5000 durante 30 min, anti-cabra (1:3000) o anti-ratón (1:5000) y se lavaron con TBS-T durante 30 min. Las membranas se visualizaron usando el kit de quimioluminiscencia Immun-Star WesternC™ (Bio-Rad) y las imágenes se capturaron usando un sistema ChemiDoc™ XRS+ y se procesaron usando el software ImageLab (Bio-Rad).

10 15 **ELISA.** Se recogieron medios de cultivo sin suero (OPTI-MEM) y se congelaron a -80 °C. La secreción de IL-1ra humana en medio de cultivo se cuantificó mediante un ensayo de inmunoadsorción ligado a enzimas (ELISA), de acuerdo con los protocolos del fabricante (R&D Systems, N.º de Cat. DY280). La curva patrón se preparó diluyendo IL-1ra humano recombinante en OPTI-MEM y la IL-1ra en medio de cultivo se midió sin diluir. Las muestras se concentraron aproximadamente 8 veces mediante centrifugación a través de filtros MWCO de 3 kDa durante 20 minutos (Amicon Ultra, N.º de Cat UFC500396). Los valores informados fueron corregidos por el factor de concentración para cada muestra.

15 20 La densidad óptica se midió a 450 nm, con una corrección de longitud de onda a 540 nm. Cada patrón y muestra se ensayó por duplicado. Las lecturas duplicadas se promediaron y se normalizaron restando la densidad óptica convencional cero promedio. Se generó una curva patrón transformando logarítmicamente los datos y realizando una

regresión lineal de la concentración de IL-1ra frente a la densidad óptica. Los valores informados son la media y el error típico de la media de tres experimentos independientes ($n = 3$) que se realizaron en días diferentes con duplicados técnicos que se promediaron para cada experimento.

- 5 **qRT-PCR.** El ARN total se aisló usando el kit de aislamiento de ARN RNeasy Plus (Qiagen). La síntesis de ADNc se realizó usando el kit de síntesis de ADNc Superscript® VILO™ (Invitrogen). La PCR en tiempo real usando PerfeCTa® SYBR® Green FastMix se realizó con el sistema de detección de PCR en tiempo real CFX96 (Bio-Rad) con cebadores oligonucleotídicos indicados en la Tabla 3 que se diseñaron usando el software Primer3Plus y se compraron en IDT.

10 **Tabla 3** Secuencias de cebadores usados para qRT-PCR.

Diana	Cebador directo	SEQ ID NO	Cebador inverso	SEQ ID NO
<i>hASCL1</i>	GGAGCTCTCGACTTCACCA	41	AACGCCACTGACAAGAAAGC	53
<i>NANOG</i>	GATTGTGGGCCTGAAGAAA	42	CAGATCCATGGAGGAAGGAA	54
<i>VEGFA</i>	AAGGAGGAGGGCAGAACATCAT	43	GGGTACTCCTGGAAGATGTCC	55
<i>TERT</i>	AAACCTTCCTCAAGTATGCC	44	GTTTGCGACCCATGTTCTC	56
<i>IL1B</i>	AGCTGATGGCCCTAAACAGA	45	AAGCCCTTGCTGTAGTGGTG	57
<i>ILR2</i>	CAGGAGGACTCTGGCACCTA	46	CGGCAGGAAAGCATCTGTAT	58
<i>IL1RN</i>	GGAACTCATGGAGGGAAAGAT	47	TGTTCTCGCTCAGGTCACTG	59
<i>HBG1/2</i>	GCTGAGTGAAC TGCACTGTGA	48	GAATTCTTGCCGAAATGGA	60
<i>MYOD1</i>	CTCTCTGCTCCTTGGCACA	49	GTGCTCTCGGGTTTCAGGA	61
<i>GAPDH</i>	CAATGACCCCTTCATTGACC	50	TTGATTGGAGGGATCTCG	62
<i>mASCL1</i>	GGAAACAAGAGCTGCTGGACT	51	GTTCCTCTGCCTCCCCATT	63
<i>mGAPDH</i>	AACTTGGCATTGTGGAAGG	52	GGATGCAAGGGATGATGTTCT	64

- 15 La especificidad del cebador se confirmó mediante electroforesis en gel de agarosa y análisis de la curva de fusión. Se calcularon las eficiencias de reacción en el intervalo dinámico apropiado para garantizar la linealidad de la curva patrón (**Fig. 10**). Los resultados se expresan como expresión de ARNm multiplicada por el gen de interés normalizado a la expresión de GAPDH mediante el método $\Delta\Delta C_T$. Los valores informados son la media y el error típico de la media de tres experimentos independientes realizados en días diferentes ($n = 3$) con duplicados técnicos que se promediaron para cada experimento.

- 20 **Sec. de ARN.** Se construyeron bibliotecas de sec. de ARN. Brevemente, la primera cadena de ADNc se sintetizó a partir de ARNm capturado con oligo dT Dynabead® (Invitrogen) usando el kit de síntesis de ADNc SuperScript® VILO™ (Invitrogen). La segunda cadena de ADNc se sintetizó usando ADN polimerasa I (New England Biolabs). El ADNc se purificó usando perlas Agencourt AMPure XP (Beckman Coulter) y Ncxtrra transposase (Illumina; 5 min a 55 °C) para fragmentar e insertar simultáneamente cebadores de secuenciación en el ADNc bicatenario. Las reacciones de transposición se detuvieron usando tampón QG (Qiagen) y el ADNc fragmentado se purificó en perlas AMPure XP. Se generaron bibliotecas de secuenciación indexadas mediante 6 ciclos de PCR.

- 25 Las bibliotecas se secuenciaron usando lecturas de un solo extremo de 50 pb en dos carriles de un instrumento Illumina HiSeq 2000, generando entre 29 millones y 74 millones de lecturas por biblioteca. Las lecturas se alinearon con transcripciones de Sec. Ref. humanas usando Bowtie (Langmead et al. *Genome biology* 10, R25 (2009)). La significancia estadística de la expresión diferencial, incluyendo corrección para pruebas de hipótesis múltiples, se calculó usando DESeq (Anders et al. *Genome biology* 11, R106 (2010)). Las lecturas de sec. de ARN sin procesar y el número de lecturas alineadas con cada transcripción de RefSeq se han depositado para acceso público en Gene Expression Omnibus (GEO), con número de registro actualmente en trámite.

- 30 **Tinción por inmunofluorescencia.** Para la detección de la expresión de Tuj1 y MAP2, las MEF transfectadas se fijaron el día 10 de cultivo en medio N3 con PFA al 4 % (EMS, Hatfield, PA) a temperatura ambiente (TA) durante 20 min. Las células se incubaron después en un tampón de bloqueo que contenía Triton X-100 al 0,2 %, BSA al 3 % p/v y suero de cabra al 10 % (Sigma-Aldrich, Saint Louis, MO) durante dos horas a temperatura ambiente con conejo anti-Tuj 1 (Covance, Princeton, Nueva Jersey, clon TUJ1 1-15-79, 1:500) y anti-MAP2 de ratón (BD, clon Ap20, 1:500), o durante 24 h adicionales a 4 °C con anti-Ascl1 de ratón (clon BD 24B72D11.1, 1:100). Las células se lavaron después tres veces con PBS, se incubaron durante 1 h a temperatura ambiente en tampón de bloqueo con cabra Alexa Fluor 488 anti-IgG de ratón y cabra Alexa Fluor 594 anti-IgG conejo (Invitrogen, 1:200) y se lavaron tres veces con PBS. Las MEF teñidas se escanearon con un microscopio de fluorescencia invertida Nikon Eclipse TE2000-U con una fase motorizada ProScanil (Prior Scientific, Rockland, MA) para producir grandes imágenes en mosaico de cada área de cultivo completa. Estos mosaicos se procesaron con un macro FIJI para establecer un umbral automático y uniforme

de cada imagen según el contraste local, excluyendo residuos pequeños y para contar el número de células Tuj1⁺ en cada pocillo.

Estadísticas. El análisis estadístico se realizó mediante la prueba de Tukey con alfa igual a 0,05 en JMP 10 Pro.

5

Ejemplo 2

Resultados

- 10 Para crear un sistema de activación transcripcional basado en CRISPR/Cas9, los restos catalíticos de Cas9 (D10A, H840A) se mutaron para crear iCas9 y se fusionaron genéticamente con un dominio de transactivación ácido VP64 C-terminal. (**Fig. 1a,b**). Se observó una expresión sólida de iCas9-VP64 a partir del plásmido transfectado en células 293T de riñón embrionario humano (HEK) mediante transferencia Western de la etiqueta del epítopo Flag N-terminal. (**Fig. 3**). El sistema CRISPR reconoce su diana mediante el emparejamiento de bases de una secuencia de 20 pb en el ARNg con una diana de ADN complementario, que es seguida por la secuencia del motivo adyacente al protoespaciador (PAM) NGG, donde N es cualquier par de bases. Las combinaciones de factores de transcripción sintéticos dirigidos a promotores humanos endógenos dan como resultado una activación robusta y sinérgica de la expresión génica. Por lo tanto, se identificaron cuatro sitios diana de ARNg seguidos de la secuencia NGG PAM en el promotor del gen ***IL1RN*** dentro de 500 pb del sitio de inicio de la transcripción (**Fig. 4, Tabla 2**). Para comparar estrategias de focalización basadas en ARNcr y ARNg, las cuatro secuencias del sitio diana se introdujeron en plásmidos¹⁷ de expresión de ARNcr y ARNg y cotransfектado con el plásmido de expresión iCas9-VP64 en células HEK293T. Aunque la inducción sustancial de la expresión de ***IL1RN*** se observó mediante qRT-PCR en muestras tratadas con la combinación de ARNcr, se lograron niveles mucho más altos con la combinación de ARNg (**Fig. 1c**). No se observaron cambios en la expresión génica en células tratadas con ARNg y un plásmido de expresión para iCas9 sin VP64, demostrando el papel crítico del dominio de activación en la modulación de la expresión génica (**Fig. 1c**). Se confirmó que la actividad nucleasa en estos sitios diana se había anulado en el sistema iCas9-VP64 mediante la realización del ensayo de inspección para detectar eventos de reparación del ADN en muestras tratadas con iCas9-VP64 y Cas9 de tipo silvestre (**Fig. 5**). Al transfectar cada uno de los cuatro ARNg individualmente o en combinación, dirigirse a múltiples sitios en el promotor con combinaciones de ARNg mostró aumentos sólidos en la expresión génica (**Fig. 1d**). Se observaron altos niveles de expresión de ***IL1RN*** solo cuando las combinaciones de ARNg se cotransfектaron con iCas9-VP64 (**Fig. 1d**), como se ve con otras clases de factores de transcripción modificados por ingeniería genética. De forma similar, la producción de la proteína antagonista del receptor de IL-1 (IL-1ra), codificada por el gen ***IL1RN***, solo se observó en tres de las seis muestras tratadas con la combinación de ARNg en tres experimentos diferentes, mientras que nunca se detectó en muestras tratadas con ARNg individuales o plásmido control (**Fig. 1e**). Para examinar la especificidad de la activación génica por iCas9-VP64, se evaluó la expresión génica global de células HEK293T tratadas con la combinación de cuatro ARNg mediante sec. de ARN. (**Fig. 1f**). Notablemente, los únicos genes con una expresión significativamente mayor en relación con el control (tasa de descubrimiento falso $\leq 3 \times 10^{-4}$) fueron las cuatro isoformas expresadas a partir del locus ***IL1RN*** (**Fig. 4**), lo que indica un alto nivel de especificidad de la activación génica.
- 40 Para demostrar la aplicabilidad general de este sistema, se diseñaron cuatro ARNg para dirigirse a cada uno de los promotores de otros ocho genes relevantes para la medicina y la biotecnología, incluyendo **ASCL1**, **NANOG**, **HBG1/2**, **MYOD1**, **VEGFA**, **TERT**, **IL1B** e **IL1R2** (**Fig. 4, Tabla 2**). La expresión forzada de **ASCL1** y **MYOD1** conduce a la transdiferenciación de varios tipos celulares en fenotipos neuronales y miogénicos, respectivamente. **NANOG** es un marcador de pluripotencia y que también se usa en estrategias de reprogramación genética. La activación de los homólogos **HBG1** y **HBG2**, que codifican la γ -globina durante el desarrollo fetal, puede usarse como estrategia terapéutica para compensar las mutaciones de la β -globina en la anemia de células falciformes. La regulación positiva de **VEGFA** mediante factores de transcripción sintéticos se ha explorado como una estrategia para mejorar la regeneración de tejidos y la cicatrización de heridas. La expresión forzada de la telomerasa, codificada por el gen **TERT**, puede usarse para inmortalizar líneas celulares. La **IL1B** codifica la citocina IL-1 β que media la inflamación y la autoinmunidad. La señalización de IL-1 β puede bloquearse mediante la expresión de IL-1ra o el receptor señuelo codificado por **IL1R2**. La expresión de cada uno de estos genes se mejoró mediante la cotransfección de plásmidos de expresión para iCas9-VP64 y los cuatro ARNg en células HEK293T, según lo determinado por qRT-PCR (**Fig. 2**). En algunos casos la expresión de un ARNg individual fue suficiente para inducir la expresión génica, pero en todos los casos la cotransfección de los cuatro ARNg condujo a efectos sinérgicos (**Fig. 2a-d**). Notablemente, la accesibilidad a la cromatina, según lo determinado por sec. de DNasa, no fue un predictor de activación génica exitosa (**Fig. 4**). La sec. de ARN se realizó en células transfectadas con iCas9-VP64 y los cuatro ARNg dirigidos a **HBG1**, tres de los cuales también se dirigían a **HBG2**. Esto reveló aumentos específicos y reproducibles en la expresión de ambos **HBG1** y **HBG2**, que no puede distinguirse por sec. de ARN, aunque no se alcanzó significación estadística debido a los bajos niveles de expresión total (**Fig. 6**). Se detectaron aumentos en la expresión de proteínas de Ascl1 y γ -globina después del tratamiento con iCas9-VP64 y los cuatro ARNg mediante transferencia Western (**Fig. 7**), corroborando niveles más altos de ARNm observados por qRT-PCR (**Fig. 2**). Se detectaron niveles basales bajos de expresión de proteínas Ascl1 y γ -globina en controles de vectores vacíos. Como evidencia preliminar de que la activación de dianas genéticas por iCas9-VP64 puede conducir a cambios secundarios en las redes génicas y los fenotipos celulares, los plásmidos de expresión para iCas9-VP64 y los cuatro ARNg dirigidos **ASCL1** fueron cotransfектados en fibroblastos embrionarios murinos (MEF) (**Fig. 8**). Se ha demostrado que la expresión forzada de Ascl1 en MEF activa parcialmente la red de
- 45
- 50
- 55
- 60
- 65

genes neuronales, incluyendo la diana en dirección 3' Tuj1. Debido a que los sitios diana del ARNg se conservan en los promotores de *ASCL1* humano y de ratón (Fig. 8a), la activación de la expresión de *ASCL1* también se observó en MEF tratados con plásmidos que codifican iCas9-VP64 y los cuatro ARNg (Fig. 8b). Adicionalmente, las células que expresan Ascl1 y el marcador neuronal Tuj1 se detectaron fácilmente mediante tinción por inmunofluorescencia 5 12 días después de la transfección en las muestras tratadas con iCas9-VP64/ARNg (Fig. 8c-h). No se observaron células positivas para Tuj1 en las células tratadas con el plásmido de control.

Hasta el momento no se ha realizado ningún estudio exhaustivo sobre la especificidad de la actividad Cas9/CRISPR en células de mamíferos. Usando sec. de ARN, se demostró que la activación génica dirigida es exquisitamente 10 específica sin activación detectable de genes inespecífica (Fig. 1f, Fig. 6). *IL1RN* y *HBG1/2* fueron elegidos para este análisis de especificidad ya que los productos génicos, IL-1ra y γ -globina, pueden no generar efectos secundarios sobre la expresión génica en células HEK293T. Explotando la actividad sinérgica de múltiples activadores transcripcionales débiles, en contraste con el uso de un único activador fuerte, puede aumentar la regulación genética 15 específica, ya que es poco probable que existan múltiples sitios adyacentes inespecífico en otro locus. De forma interesante, el gen *IL32* estaba moderadamente regulado negativamente (tasa de descubrimiento falso <0,03) tanto en las muestras tratadas con iCas9-VP64 como en las *IL1RN*- o ARNg dirigidos a *HBG1/2* en comparación con muestras control tratadas solo con un plásmido de expresión vacío (Fig. 1f, Fig. 6). Debido a que ambas muestras dirigidas a *IL1RN* y *HBG1/2* se vieron afectadas de manera similar, es poco probable que esto sea el resultado de una actividad iCas9-VP64 inespecífico relacionada con la identidad de las secuencias diana.

20 Para evaluar la especificidad con la que iCas9-VP64 se une al genoma, la secuenciación ChIP se realizó usando un anticuerpo anti-HA en células tratadas con iCas9-VP64 y cuatro ARNg dirigidos al promotor *IL1RN*. El experimento reveló que iCas9 se dirige al promotor *IL1RN* (Fig. 15). Por otra parte, el experimento reveló un nivel extremadamente alto de especificidad. La iCas9 tenía solo 10 sitios de unión inespecífica potenciales (FDR <5 %). Para consultar más 25 a fondo la especificidad, Se realizaron experimentos de secuenciación de ARN con EGEM iCas9 y se encontró que solo las isoformas del gen *IL1RN* aumentaron en expresión en relación con el control (FDR $\leq 3 \times 10^{-4}$).

Ejemplo 3

30 CRISPR dirigidos al gen de la distrofina: métodos y materiales

Construcciones de plásmidos. Se usaron los casetes de expresión para el ARN_{gm} de *S. pyogenes* y la nucleasa Cas9 (hCas9) optimizada con codones humanos, como se ha descrito anteriormente (Perez-Pinera *et al.*, Nat Methods 10:973-976 (2013)). Para crear un sistema indicador fluorescente para enriquecer las células modificadas con 35 CRISPR/Cas9, se sintetizó un GeneBlock (IDT) que contenía una porción del extremo 3' de la secuencia codificante de Cas9 fusionada a un péptido de omisión T2A inmediatamente en dirección 5' de un sitio de clonación múltiple y posteriormente se clonó en el vector de expresión hCas9. Después se clonó un gen indicador eGFP en el vector T2A para permitir la cotraducción de las proteínas Cas9 y eGFP del mismo vector de expresión (hCas9-T2A-GFP, SEQ ID NO: 116).

40 **Cultivo celular y transfección.** Las células HEK293T se obtuvieron del American Tissue Collection Center (ATCC) a través del Duke Cell Culture Facility y se mantuvieron en DMEM suplementado con suero de ternera bovino al 10 % de y penicilina/estreptomicina al 1 %. Los mioblastos inmortalizados (Mamchaoui, K. *et al.* Skelet Muscle 1, 1-11 (2011)) (uno de un donante natural y dos líneas derivadas de pacientes con DMD $\Delta 48-50$) se mantuvieron en medio 45 de músculo esquelético (PromoCell) suplementado con suero de ternera bovino al 20 % (Sigma), fetuina 50 μ g/ml, factor de crecimiento epidérmico humano 10 ng/ml (Sigma), factor de crecimiento de fibroblastos básico humano 1 ng/ml (Sigma), insulina humana 10 μ g/ml (Sigma), GlutaMAX al 1 % (Invitrogen) y penicilina/estreptomicina al 1 % (Invitrogen). Los fibroblastos dérmicos primarios de DMD se obtuvieron del depósito de células Coriell (GM05162A, $\Delta 46-50$) y se mantuvieron en DMEM suplementado con suero bovino fetal al 10 %, factor de crecimiento de fibroblastos 50 básico humano 1 ng/ml y penicilina/estreptomicina al 1 %. Todas las líneas celulares se mantuvieron a 37 °C y CO₂ al 5 %.

Las células HEK293T se transfecaron con Lipofectamine 2000 (Invitrogen) con 400 ng de cada vector de expresión 55 según el protocolo del fabricante en placas de 24 pocillos. Se transfecaron mioblastos inmortalizados y fibroblastos primarios con 5 μ g de cada vector de expresión mediante electroporación usando Gene Pulser XCell (BioRad) con PBS como tampón de electroporación usando condiciones optimizadas para cada línea (Fig. 1) (Ousterout *et al.* Mol Ther 21:1718-1726 (2013)). Las eficiencias de transfección se midieron mediante el suministro de un plásmido de expresión de eGFP (pmaxGFP, Clontech) y el uso de citometría de flujo. Estas eficiencias fueron habitualmente ≥ 95 % para HEK293T y ≥ 70 % para los fibroblastos primarios y los mioblastos inmortalizados. Para todos los experimentos, 60 la masa indicada de plásmido electroporado corresponde a la cantidad usada para cada sistema basado en CRISPR/Cas9.

Cuantificación Cel-1 de modificación génica endógena (Ensayo de inspección).

Las lesiones inducidas por el sistema basado en CRISPR/Cas9 en el sitio diana endógeno se cuantificaron usando el 65 ensayo de nucleasa de inspección (Guschin, D.Y. *et al.* Meth Mol Biol 649, 247-256 (2010)), que puede detectar mutaciones características de NHEJ mediada por nucleasas. Después de la transfección, las células se incubaron

durante 3 o 10 días a 37 °C y se extrajo el ADN genómico usando el kit DNeasy Blood and Tissue (QIAGEN). El locus diana se amplificó mediante 35 ciclos de PCR con el kit AccuPrime High Fidelity PCR (Invitrogen) usando cebadores específicos para cada locus (véase la **Tabla 4**), tales como 5'-GAGTTGGCTCAAATTGTTACTCTT-3' (SEQ ID NO: 60) y 5'-GGGAAATGGCTAGGAGAGTAAAGT-3' (SEQ ID NO: 61).

5

Tabla 4 Resumen de los 10 principales sitios inespecíficos predichos *in silico* y actividad en cada sitio detectada por el ensayo de inspección en células HEK293T transfectadas con Cas9 y los casetes de expresión de ARNgm indicados. n.d.: no detectado.

SEQ ID NO.	Guía	Diana	Secuencia	PAM	Puntuación	Cr	Gen	Intrón/Exón	N.º de MM	% de indeles
67	CR3	Guía	GCCTACTCAGACTGTTACTC	-	-	-	-	-	-	-
150		Diana	tCCTACTCAGACTGTTACTC	TGG	-	X	DMD	Exón	1	13,0
151		OT1	tCCTACTCAcACTGTTACTC	AGG	7,4	1	STRIP1	Intrón	2	9,3
152		OT2	aCCTgCTCAcACTGTTACTC	CAG	2,5	2	ARHGAP25	Intrón	3	n.d.
153		OT3	GCaTtCTCAaACTGTTACTC	AGG	2,4	13	Ninguno	Ninguno	3	n.d.
154		OT4	GgaTtCTCAcACTGTTACTC	GGG	13	14	PGPEPI	Exón	4	n.d.
155		OT5	aCaTACTtAtACTGTTACTC	TAG	1,3	19	MDGA2	Intrón	4	n.d.
156		OT6	tatTcCTaAGACTGTTACTC	AAG	0,9	8	LPPR1	Intrón	5	n.d.
157		OT7	aaggACTaAGACTGTTACTC	GGG	0,9	9	RXF122	Intrón	5	n.d.
158		OT8	GagctCTCAtACTGTTACTC	TAG	0,8	3	DNMBP	Exón	5	n.d.
159	CR1	OT9	GCaaAaTgAGACTGTTACTC	CAG	0,8	5	SLC12A2	Intrón	4	n.d.
160		OT10	cCtAttCAGACTGTTACTC	AAG	0,8	4	KCNIP4	Intrón	4	n.d.
65		Guía	GATTGGCTTGATTTCCCTA	-	-	-	-	-	-	-
161		Diana	cATTGGCTTGATTTCCCTA	GGG	-	X	DMD	Intrón	1	8,3
162		OT1	aATTGGCATTGATTTCCCTA	GAG	7,1	16	Ninguno	Ninguno	2	0,8
163		OT2	cATTGGCTTaATTTCCCTA	TAG	4,8	4	Ninguno	Ninguno	2	n.d.
164		OT3	GATaGGCTgTGATTTCCCTA	GAG	3,9	9	Ninguno	Ninguno	2	n.d.
165		OT4	GAaTaGCcTTGATTTCCCTA	AAG	2,4	1	Ninguno	Ninguno	3	n.d.
166		OT5	aATTtGTTTGATTTCCCTg	AGG	1,5	1	TIMM17A	Intrón	3	n.d.
167		OT6	GATgtGCTTGATTTCCCTt	GGG	1,4	17	MYO1D	Intrón	3	n.d.
168		OT7	aATTGGtTTTaATTTCCCTA	AAG	1,1	8	PIK1A	Intrón	3	n.d.
169		OT8	aATTGGgTTTGATTTCCCTt	TGG	1,1	11	MS4A1	Intrón	3	n.d.
170	CR5	OT9	GATgGGTTTATTTCCCTA	GAG	1,0	11	Ninguno	Ninguno	3	n.d.
171		OT10	GAaTGGTTTGATTTCCCTg	GAG	1,0	11	Ninguno	Ninguno	3	n.d.
69		Guía	GCAGTTGCCATAAGAACTGGT	-	-	-	-	-	-	-
172		Diana	aCAGTTGCCATAAGAACTGGT	GGG	-	X	DMD	Intrón	1	14,0
173		OT1	cCAGTTGtCTAAGAACTGGg	GAG	1,5	5	NRG1	Intrón	3	n.d.
174		OT2	GCAGTTGCCGtGAACGGT	AGG	1,4	X	Ninguno	Ninguno	2	n.d.
175		OT3	GCAGaTGCAGaAAGAACTGGT	GAG	1,4	19	SMIM7	Intrón	3	n.d.
176		OT4	GCAGTTcCAGaAAGAACTGGT	GAG	0,9	11	GLB1L2	Intrón	3	n.d.
177		OT5	caAcTTGCCCTaGAACGGT	AGG	0,7	8	ASAP1	Intrón	4	n.d.
178		OT6	aCAccTGCCTAAGAACTGGa	GGG	0,7	11	Ninguno	Ninguno	4	n.d.
179		OT7	tCAGgTGgCTAAGAACTGGg	TGG	0,7	14	NIN	Intrón	4	n.d.
180		OT8	GaAGTTGgCcAAGAACTGGa	GAG	0,6	7	Ninguno	Ninguno	4	n.d.
181		OT9	GCtGcTGCCcAAGAACTGGc	AGG	0,6	11	AMOTLI	Intrón	4	n.d.

182		OT10	tCAGeTGgCTAAGAACgGGT	AAG	0,6	7	ACTR3C	Intrón	4	n.d.
70	CR6	Guía	GGGGCTCCACCCCTACGAGT	-	-	-	-	-	-	-
183		Diana	aGGGCTCCACCCCTACGAGT	GGG	-	X	DMD	Intrón	1	19,9
184		OT1	GcaGCTCAGCCCTACGAGT	CAG	0,8	3	Ninguno	Ninguno	4	n.d.
185		OT2	GGGGCTCAgCaTCACGAGT	GAG	0,8	8	Ninguno	Ninguno	3	n.d.
186		OT3	GGGGCTCtCCCTCACTAGT	GAG	0,6	8	Ninguno	Ninguno	3	n.d.
187		OT4	GGGGatCCACCACTCACcAGT	CAG	0,6	2	Ninguno	Ninguno	3	n.d.
188		OT5	aGGGCTggACCCTCACaAGT	AAG	0,4	16	AXIN1	Intrón	4	n.d.
189		OT6	tGGtCTCcCCCCeCACGAGT	GGG	0,4	2	Ninguno	Ninguno	4	n.d.
190		OT7	aGGGCTCCeCACGAGT	GAG	0,3	5	Ninguno	Ninguno	4	n.d.
191		OT8	GaGGCTCCAaCTCACcAGT	GAG	0,3	11	Ninguno	Ninguno	4	n.d.
192		OT9	GGaGCTgCcCCtTCACGAGT	GGG	0,3	3	Ninguno	Ninguno	4	n.d.
193		OT10	atGaCTCCCCTCAaGAGT	AAG	0,3	8	AGPAT5	Ninguno	4	n.d.
100	CR36	Guía	GCCTTCTTATCCCCATCG	-	-	-	-	-	-	-
194		Diana	GCCTTCTTATCCCCATCG	AGG	-	X	DMD	Intrón	0	20,6
195		OT1	GtCTgCTgTgTCCCCTATCG	GGG	1,3	21	Ninguno	Ninguno	4	n.d.
196		OT2	cCCTCTCtTATCCCCCTgTCG	TGG	13	8	Ninguno	Ninguno	3	n.d.
197		OT3	GCCTTCTTATCCCCCTtTCT	TGG	0,9	10	Ninguno	Ninguno	2	0,5
198		OT4	GCgcTCTTTtTCCCCTATCt	TAG	0,6	16	Ninguno	Ninguno	4	n.d.
199		OT5	GCCcTCTgTcTCCCCTgTCG	CAG	0,5	1	NFASC	Ninguno	4	n.d.
200		OT6	tCCATCTtTgTCCCCTATgG	AGG	0,5	10	Ninguno	Ninguno	4	n.d.
201		OT7	aCCtTCTCTcTCCCCTATAg	AGG	0,5	5	LOC100996185	Intrón	4	n.d.
202		OT8	GtTTCTTTtTCCCCTATgG	GAG	0,5	3	Ninguno	Ninguno	4	n.d.
203		OT9	tgCTTCTTaATCCCCCTATCa	AAG	0,4	7	Ninguno	Ninguno	4	n.d.
204		OT10	aCCTTCTTAcTCCCCTATCc	GGG	0,4	10	ADARB2	Ninguno	4	n.d.

Los productos de PCR resultantes se fundieron aleatoriamente y se volvieron a hibridar en un termociclo con el programa: 95 °C durante 240 s, seguido de 85 °C durante 60 s, 75 °C durante 60 s, 65 °C durante 60 s, 55 °C durante 60 s, 45 °C durante 60 s, 35 °C durante 60 s y 25 °C durante 60 s con una velocidad de -0,3 °C/s entre etapas.

5 Despues de la rehibridación, se mezclaron 8 µl de producto de PCR con 1 µl de Nucleasa S de inspección y 1 µl de Potenciador S (Transgenomic) y se incubaron a 42 °C durante 1 h. Despues de la incubación, se cargaron 6 µl del producto de digestión en un gel de poliacrilamida TBE al 10 % y se ejecutaron a 200 V durante 30 min. Los geles se tiñeron con bromuro de etidio y se cuantificaron usando ImageLab (Bio-Rad) mediante densitometría como se describió anteriormente (Guschin, *et al.* *Meth Mol Biol* 649, 247-256 (2010)).

10 **Clasificación de mioblastos de células activada por fluorescencia.** Los mioblastos DMD se sometieron a electroporación con 5 microgramos de cada uno de los vectores de expresión hCas9-T2A-GFP y ARNgm y se incubaron a 37 °C y CO₂ al 5 %. Tres días después de la electroporación, las células se tripsinizaron y se recogieron para su clasificación FACS usando una máquina clasificadora FACSvantage II. Se recogieron y cultivaron células positivas para GFP para su análisis.

15 **Ensayo basado en PCR para detectar eliminaciones genómicas.** Los loci del exón 51 o del exón 45-55 se amplificaron a partir de ADN genómico mediante PCR (kit de PCR de alta fidelidad Invitrogen AccuPrime) usando cebadores que flanquean cada locus. Los cebadores flanqueantes fueron Cell-CR1/2-F y Cell-CR5-R para el exón 51 o Cell-CR6-F y Cell-CR36-R para el análisis de los exones 45-55 (Tabla 4). Los productos de la PCR se separaron en geles de TAE-agarosa y se tiñeron con bromuro de etidio para su análisis.

20 **Detección de translocaciones basada en PCR.** Los loci con posibles translocaciones previstas se amplificaron mediante una PCR anidada de dos pasos (kit de PCR Invitrogen AccuPrime High Fidelity para cada etapa) de ADN genómico de células transfectadas con Cas9 solo (control) o Cas9 con ARNgm. En la primera etapa, las translocaciones que pueden producirse en cada sitio diana de ARNgm específico e inespecífico se amplificaron mediante 35 ciclos de PCR usando combinaciones de cebadores de inspección para cada locus que se modificaron para incluir sitios de restricción para facilitar el análisis de clonación y la secuenciación (Tabla 4). Un microlitro de cada

reacción de PCR se sometió a una segunda ronda de amplificación mediante 35 rondas de PCR usando conjuntos de cebadores anidados diseñados a medida para cada translocación individual prevista (Tabla 4). Cada segundo cebador de PCR anidado se une dentro de la misma región aproximada dentro del amplicón primario; sin embargo, cada par se optimizó usando el software de bioinformática en línea Primer3 para garantizar la detección específica de cada

5 translocación. Los amplicones de PCR correspondientes a la longitud esperada de las translocaciones predichas y solo presentes en células tratadas con ARN_m se purificaron (kit de extracción en gel QIAGEN) y se analizaron mediante secuenciación de Sanger.

10 **Análisis de ARNm.** Los mioblastos inmortalizados se diferenciaron en miofibras reemplazando el medio de crecimiento por DMEM suplementado con insulina-transferrina-selenio al 1% (Invitrogen N.º 51500056) y penicilina/estreptomicina al 1% (Invitrogen N.º 15140) durante 5 días antes de que las células se tripsinizaran y se recolectaran. El ARN total se aisló de estas células usando el mini kit RNeasy Plus (QIAGEN) de acuerdo con las instrucciones del fabricante. El ARN se retrotranscribió a ADNc usando el kit de síntesis de ADNc VILO (Life Technologies N.º 11754) y 1,5 microgramos de ARN durante 2 h a 42 °C de acuerdo con las instrucciones del fabricante. Los loci diana se amplificaron mediante 35 ciclos de PCR con el kit AccuPrime High Fidelity PCR (Invitrogen) usando cebadores que se hibridan con los exones 44 y 52 para detectar la eliminación del exón 51 por CR1/5 o CR2/5 o cebadores que se hibridan con los exones 44 y 60 para detectar la eliminación del exón 45-55 por CR6/36 (Tabla 4). Los productos de la PCR se hicieron correr en geles de TAE-agarosa y se tiñeron con bromuro de etidio para su análisis. Las bandas de PCR resueltas se clonaron y analizaron mediante secuenciación de Sanger para verificar las uniones de exones esperadas. La Tabla 5 enumera las secuencias de los cebadores usados en el Ejemplo 4.

Tabla 5

SEQ ID NO.	Nombre del cebador	Secuencia del cebador	Notas
205	Cell-CR1/2-F	GAGAGGTTATGTGGCTTACCA	Cebador de inspección directo para CR1/2
206	Cell-CR1-R	AAAAATGCTTCCCACTTGC	Cebador de inspección inverso para CR1
207	Cell-CR2-R	CTCATTCTCATGCCTGGACA	Cebador de inspección inverso para CR2
208	Cell-CR3-F	GAGTTGGCTCAAATTGTTACTCTT	Cebador de inspección directo para CR3
209	Cell-CR3-R	GGGAAATGGTCTAGGAGAGTAAAGT	Cebador de inspección inverso para CR3
210	Cell-CR4/31-F	GTGGCTCAAATTGTTACTCTTC	Cebador de inspección directo para CR4 o CR31
211	Cell-CR4/31-R	GTGAGAGTAATGTGTTGCTGAGAG	Cebador de inspección inverso para CR4 o CR31
212	Cell-CR5-F	CGGGCTTGGACAGAACCTAC	Cebador de inspección directo para CR5
213	Cell-CR5-R	CTGCGTAGTGCCAAAACAAA	Cebador de inspección inverso para CR5
214	Cell-CR6-F	TAATTCATTGAAGAGTGGCTGAA	Cebador de inspección directo para CR6
215	Cell-CR6-R	AAGCCCTGTGGTAGTCAGT	Cebador de inspección inverso para CR6
216	Cell-CR7-F	TGAGTCATGTTGGATAACCAGTCT	Cebador de inspección directo para CR7
217	Cell-CR7-R	GAAGGTCAGGAACATACAATTCAA	Cebador de inspección inverso para CR7
218	Cell-CR10/II-F	GATATGGGCATGTCAGTTCATAG	Cebador de inspección directo para CR10 o CR11
219	Cell-CR10/II-R	TGCTGTTGATTAATGGTTGATAGG	Cebador de inspección inverso para CR10 o CR11
220	Cell-CR12/13-F	TTTAAATTGCCATGTTGTGTC	Cebador de inspección directo para CR12 o CR13
221	Cell-CR12/13-R	ATGAATAACCTAATGGGCAGAAAA	Cebador de inspección inverso para CR12 o CRB

222	Cell-CR14/15-F	TCAAGTCGCTTCATTTGATAGAC	Cebador de inspección directo para CR14 o CR15
223	Cell-CR14/15-R	CACAACAAAACATATAGCCAAAGC	Cebador de inspección inverso para CR14 o CR15
224	Cell-CR16/17-F	TGCTGCTAAAATAACACAAATCAGT	Cebador de inspección directo para CR16 o CR17
225	Cell-CR16/17-R	CTGTGCCTATTGTGGTTATCCTG	Cebador de inspección inverso para CR16 o CR17
226	Cell-CR18/19-F	ATTGATCTGCAATAACATGTGGAGT	Cebador de inspección directo para CR18 o CR19
227	Cell-CR18/19-R	TTTG CCTCTGCTATTACAGTATGG	Cebador de inspección inverso para CR18 o CR19
228	Cell-CR20/21-F	TGTAGGGTGGTTGGCTAAAATAAT	Cebador de inspección directo para CR20 o CR21
229	Cell-CR20/21-R	TTTTGCACAGTCATAAACACAAA	Cebador de inspección inverso para CR20 o CR21
230	Cell-CR22/23-F	GGCTGGTCTCACAAATTGTACTTTA	Cebador de inspección directo para CR22 o CR23
231	Cell-CR22/23-R	CATTATGGACTGAAAATCTCAGCA	Cebador de inspección inverso para CR22 o CR23
232	Cell-CR24/25-F	ATCATCCTAGCCATAACACAAATGA	Cebador de inspección directo para CR24 o CR25
233	Cell-CR24/25-R	TTCAGCTTTAACGTGATTTCTGT	Cebador de inspección inverso para CR24 o CR25
234	Cell-CR26/27-F	GGATTCAAGCTGTTACGAAGT	Cebador de inspección directo para CR26 o CR27
235	Cell-CR26/27-R	TTTAGCTGGATTGGAAAAACAAAT	Cebador de inspección inverso para CR26 o CR27
236	Cell-CR28/29-F	AACTCACCCCATTGTTGGTATATT	Cebador de inspección directo para CR28 o CR29
237	Cell-CR28/29-R	CCTTGTCAAATACCGAAATACAT	Cebador de inspección inverso para CR28 o CR29
238	Cell-CR33-F	CACATAATTCATGAACCTGGCTTC	Cebador de inspección directo para CR33
239	Cell-CR33-R	TAGTAGCTGGGAGGAAGATAACAG	Cebador de inspección inverso para CR33
240	Cell-CR34-F	TTTTGTTTAATTGCGACTGTGT	Cebador de inspección directo para CR34
241	Cell-CR34-R	AGAAAAGGGTTTCTTTGACTT	Cebador de inspección inverso para CR34
242	Cell-CR35-F	CATTGTGACTGGATGAGAAGAAC	Cebador de inspección directo para CR35
243	Cell-CR35-R	AACGGCTGTTATTAAAGTCCTCAG	Cebador de inspección inverso para CR35
244	Cell-CR36-F	CAAGTCAGAAGTCAC TGCTTGT	Cebador de inspección directo para CR36
245	Cell-CR36-R	TTTATGTGCAGGAATCAGTCTGT	Cebador de inspección inverso para CR36
246	Dys-E44-F	TGGCGCGTTTCATTAT	Unión del cebador RT-PCR directo en el exón 44
247	Dys-E52-R	TTCGATCCGTAATGATTGTTCTAGCC	Unión del cebador RT-PCR inverso en el exón 52
248	Dys-E60-R	GGTCTTCCAGAGTGCTGAGG	Unión del cebador RT-PCR inverso en el exón 60
249	CR3-Cell-OT1-F	TGTGTGCTTCTGTACACATCATCT	Cebador de inspección directo para CR3 inespecífico 1

ES 2 987 399 T3

250	CR3-Cell-OT1-R	AGATTCAACCCTAAAAACTGAG	Cebador de inspección inverso para CR3 inespecífico 1
251	CR3-Cell-OT2-F	TAAACTCTTCTTTCCGCAATTC	Cebador de inspección directo para CR3 inespecífico 2
252	CR3-Cell-OT2-R	CAAGGTGACCTGCTACCTAAAAAT	Cebador de inspección inverso para CR3 inespecífico 2
253	CR3-Cell-OT3-F	TATGACCAAGGCTATGTGTTCACT	Cebador de inspección directo para CR3 inespecífico 3
254	CR3-Cell-OT3-R	ACAGCCTCTCTCCAGTAACATTCT	Cebador de inspección inverso para CR3 inespecífico 3
255	CR3-Cell-OT4-F	TATTCTGCAGTGTTTCACATT	Cebador de inspección directo para CR3 inespecífico 4
256	CR3-Cell-OT4-R	ATATTTAAGCCAAGACCCAACAA	Cebador de inspección inverso para CR3 inespecífico 4
257	CR3-Cell-OT5-F	CTTTCAACTGTCTGTCGATTGCT	Cebador de inspección directo para CR3 inespecífico 5
258	CR3-Cell-OT5-R	AACAGCCTCTTCATTGTTCTCT	Cebador de inspección inverso para CR3 inespecífico 5
259	CR3-Cell-OT6-F	CTCTGGAACTTGTCTCTGTCTTGA	Cebador de inspección directo para CR3 inespecífico 6
260	CR3-Cell-OT6-R	CTTCCTCGTTCTCATGTTACTA	Cebador de inspección inverso para CR3 inespecífico 6
261	CR3-Cell-OT7-F	CCTTATATCCGTATCGCTACTCT	Cebador de inspección directo para CR3 inespecífico 7
262	CR3-Cell-OT7-R	CATATCTGCTAACTCCGCACAC	Cebador de inspección inverso para CR3 inespecífico 7
263	CR3-Cell-OT8-F	ACAGGTGTTATGTTGCTGCATCT	Cebador de inspección directo para CR3 inespecífico 8
264	CR3-Cell-OT8-R	ACTCCATTCCCAGATTAGTTATGC	Cebador de inspección inverso para CR3 inespecífico 8
265	CR3-Cell-OT9-F	CTGTTTCTTGTGAGAGTGGAGA	Cebador de inspección directo para CR3 inespecífico 9
266	CR3-Cell-OT9-R	TGTAAGGTGGTCAAACCTGCTCTA	Cebador de inspección inverso para CR3 inespecífico 9
267	CR3-Cell-OT10-F	TTTTCCCTAGTACCCACAGATT	Cebador de inspección directo para CR3 inespecífico 10
268	CR3-Cell-OT10-R	TCCCTGATTCTCTCATTTGTGTTA	Cebador de inspección inverso para CR3 inespecífico 10
269	CR1-Cell-OT1-F	TTGGGAACATCAGAGAAAGTATGA	Cebador de inspección directo para CR1 inespecífico 1
270	CR1-Cell-OT1-R	ACAAATTACAGTCTCCTGGAAAG	Cebador de inspección inverso para CR1 inespecífico 1
271	CR1-Cell-OT2-F	AGTAGCTTACCTTGGCAGAGAAAA	Cebador de inspección directo para CR1 inespecífico 2
272	CR1-Cell-OT2-R	TGACATACTGTTACCCCTTGCAGT	Cebador de inspección inverso para CR1 inespecífico 2
273	CR1-Cell-OT3-F	GAAAGGCTCAGTGAATGTTGTT	Cebador de inspección directo para CR1 inespecífico 3
274	CR1-Cell-OT3-R	CACTGCATCATCTCATTAAATCAA	Cebador de inspección inverso para CR1 inespecífico 3
275	CR1-Cell-OT4-F	CCCATATATTGATTACCCACA	Cebador de inspección directo para CR1 inespecífico 4
276	CR1-Cell-OT4-R	TATCAGAACGAGCACTAAAAGCAC	Cebador de inspección inverso para CR1 inespecífico 4
277	CR1-Cell-OT5-F	TTGGGAGGCTGAGGTACAAG	Cebador de inspección directo para CR1 inespecífico 5

278	CR1-Cell OT5-R	GAATGAAAAACAAACAGAAGGTGA	Cebador de inspección inverso para CR1 inespecífico 5
279	CR1-Cell-OT6-F	CTCCTCATCTGTACCCTCAATCT	Cebador de inspección directo para CR1 inespecífico 6
280	CR1-Cell-OT6-R	AGAGTGGCATCTAGTGTAGTGAG	Cebador de inspección inverso para CR1 inespecífico 6
281	CR1-Cell-OT7-F	TACCAAAAGCTTCCTGTTTACC	Cebador de inspección directo para CR1 inespecífico 7
282	CR1-Cell-OT7-R	GTAAGTTGGATGGCCTATTCTTG	Cebador de inspección inverso para CR1 inespecífico 7
283	CR1-Cell-OT8-F	GAAGGAAATGCAAGGATAACAAGAT	Cebador de inspección directo para CR1 inespecífico 8
284	CR1-Cell-OT8-R	TGATTGAAAGAACATTCCAGAAA	Cebador de inspección inverso para CR1 inespecífico 8
285	CR1-Cell-OT9-F	TCAGAAGGAAAATTGAAATTGGTT	Cebador de inspección directo para CR1 inespecífico 9
286	CR1-Cell-OT9-R	CAGATGTGTTCTTCATCATTCTC	Cebador de inspección inverso para CR1 inespecífico 9
287	CR1-Cell-OT10-F	TTCTCTTAGGGAAAGCTCTCAA	Cebador de inspección directo para CR1 inespecífico 10
288	CR1-Cell-OT10-R	GGGTATAGATCATATGGAGGGAAAG	Cebador de inspección inverso para CR1 inespecífico 10
289	CR5-Cell-OT1-F	AGATGATCTGCCCACCTCAG	Cebador de inspección directo para CR5 inespecífico 1
290	CR5-Cell-OT1-R	CTTTCTTCCTCATTAGTGGCAAT	Cebador de inspección inverso para CR5 inespecífico 1
291	CR5-Cell-OT2-F	ATGAATTGCAGATTGATGGTACTG	Cebador de inspección directo para CR5 inespecífico 2
292	CR5-Cell-OT2-R	TCTCACCAAGAACCAAATTGTCTA	Cebador de inspección inverso para CR5 inespecífico 2
293	CR5-Cell-OT3-F	GTAAGGATACCTGGCAACAGTCTT	Cebador de inspección directo para CR5 inespecífico 3
294	CR5-Cell-OT3-R	TTAACGAATTGTGAGATTGCTGT	Cebador de inspección inverso para CR5 inespecífico 3
295	CR5-Cell-OT4-F	TCAGAAAGTCAAGTAGCACACACA	Cebador de inspección directo para CR5 inespecífico 4
296	CR5-Cell-OT4-R	AGAAGCACACACTCAGGTAAAGC	Cebador de inspección inverso para CR5 inespecífico 4
297	CR5-Cell-OT5-F	TCTTGGGGGAATAATGACTAAAA	Cebador de inspección directo para CR5 inespecífico 5
298	CR5-Cell-OT5-R	TTTGGCATTATGGGAATAAAACT	Cebador de inspección inverso para CR5 inespecífico 5
299	CR5-Cell-OT6-F	ACTAATTCTGGTCAAGCCCATCA	Cebador de inspección directo para CR5 inespecífico 6
300	CR5-Cell-OT6-R	TTAAGACATCGGATGAACAGAAAG	Cebador de inspección inverso para CR5 inespecífico 6
301	CR5-Cell-OT7-F	AGAAGCTTCTGACATGATCTGC	Cebador de inspección directo para CR5 inespecífico 7
302	CR5-Cell-OT7-R	TCAATTGCATTAGGACTTAGACCA	Cebador de inspección inverso para CR5 inespecífico 7
303	CR5-Cell-OT8-F	GTAAATTACCTGTGAAGCCCTTG	Cebador de inspección directo para CR5 inespecífico 8
304	CR5-Cell-OT8-R	CGGAAAACAGATCCACTTATGAT	Cebador de inspección inverso para CR5 inespecífico 8
305	CR5-Cell-OT9-F	AAATCCACTGGAAACATCTTGAGT	Cebador de inspección directo para CR5 inespecífico 9

306	CR5-Cell-OT9-R	AGTCTCTTCAGAACATGCCCTAT	Cebador de inspección inverso para CR5 inespecífico 9
307	CR5-Cell-OT10-F	GCTTGGTGGCACATAACCTGTAG	Cebador de inspección directo para CR5 inespecífico 10
308	CR5-Cell-OT10-R	GGTAGGTAGATTGCTTGCTTGT	Cebador de inspección inverso para CR5 inespecífico 10
309	CR6-Cell-OT1-F	AGCTCTCAGCAGAGTAGGGATTAA	Cebador de inspección directo para CR6 inespecífico 1
310	CR6-Cell-OT1-R	GTGAGTCTACTGCACCCCATC	Cebador de inspección inverso para CR6 inespecífico 1
311	CR6-Cell-OT2-F	TGACACTGTGAAGTCAATTCTGTC	Cebador de inspección directo para CR6 inespecífico 2
312	CR6-Cell-OT2-R	TCAAGAACTTGACAATGAGCAAAT	Cebador de inspección inverso para CR6 inespecífico 2
313	CR6-Cell-OT3-F	TATCCGATCCACTGTTGTGT	Cebador de inspección directo para CR6 inespecífico 3
314	CR6-Cell-OT3-R	CAGGAGACCCAAAACCCTCTAC	Cebador de inspección inverso para CR6 inespecífico 3
315	CR6-Cell-OT4-F	TTGTTCTACAAATAGGGCTTCCTT	Cebador de inspección directo para CR6 inespecífico 4
316	CR6-Cell-OT4-R	TGTTAAGTTGGGCTTATGTTCT	Cebador de inspección inverso para CR6 inespecífico 4
317	CR6-Cell-OT5-F	CACAAGTCTCACTGCACAAACAT	Cebador de inspección directo para CR6 inespecífico 5
318	CR6-Cell-OT5-R	TGACCCATGATTATCTCTCTTGA	Cebador de inspección inverso para CR6 inespecífico 5
319	CR6-Cell-OT6-F	TTCAGCTCTGATTGGTTTAATG	Cebador de inspección directo para CR6 inespecífico 6
320	CR6-Cell-OT6-R	CCAATTCCCTAATTTCCCTACAG	Cebador de inspección inverso para CR6 inespecífico 6
321	CR6-Cell-OT7-F	ATCTCAGACCAGGAGGGAGAC	Cebador de inspección directo para CR6 inespecífico 7
322	CR6-Cell-OT7-R	CCTCAGGGTCAGTACATTTTCAG	Cebador de inspección inverso para CR6 inespecífico 7
323	CR6-Cell-OT8-F	TTCTTAGGACATTGCTCCACATAC	Cebador de inspección directo para CR6 inespecífico 8
324	CR6-Cell-OT8-R	GCAAACATAATGCAACTCGTAATC	Cebador de inspección inverso para CR6 inespecífico 8
325	CR6-Cell-OT9-F	GCAAGGGAGTCTGTGTCTTG	Cebador de inspección directo para CR6 inespecífico 9
326	CR6-Cell-OT9-R	TCATTTAAGTGGCTGTTCTGTGTT	Cebador de inspección inverso para CR6 inespecífico 9
327	CR6-Cell-OT10-F	ACAAAACAGAGAGAAAAGGCAGAG	Cebador de inspección directo para CR6 inespecífico 10
328	CR6-Cell-OT10-R	GTTTTGATTCTGGTGCCTACAG	Cebador de inspección inverso para CR6 inespecífico 10
329	CR36-Cell-OT1-F	ACTGAAGCTGAAGCCCCAGTC	Cebador de inspección directo para CR36 inespecífico 1
330	CR36-Cell-OT1-R	ACATGAGCTCTCAGGTTCTGAC	Cebador de inspección inverso para CR36 inespecífico 1
331	CR36-Cell-OT2-F	TCAAACCTAGATGGTCCCTATGTT	Cebador de inspección directo para CR36 inespecífico 2
332	CR36-Cell-OT2-R	GTACCCCTGAAAATGTAGGGTGA	Cebador de inspección inverso para CR36 inespecífico 2
333	CR36-Cell-OT3-F	CACTTCCCAGTGAGGCAAT	Cebador de inspección directo para CR36 inespecífico 3

334	CR36-Cell-OT3-R	CTATACTGGGGCTGACTTGCTAC	Cebador de inspección inverso para CR36 inespecífico 3
335	CR36-Cell-OT4-F	TCGTATAGGTTACTTGGCTCACA	Cebador de inspección directo para CR36 inespecífico 4
336	CR36-Cell-OT4-R	AGGGATCTTACTCCTCAGTGTGT	Cebador de inspección inverso para CR36 inespecífico 4
337	CR36-Cell-OT5-F	TGTAGAAGTTGGAATATCCTGCTG	Cebador de inspección directo para CR36 inespecífico 5
338	CR36-Cell-OT5-R	GTCAACAATTGATCTCAGGCTTC	Cebador de inspección inverso para CR36 inespecífico 5
339	CR36-Cell-OT6-F	CTCAGTACTAAAGATGGACGCTG	Cebador de inspección directo para CR36 inespecífico 6
340	CR36-Cell-OT6-R	AATCATTCAGTCTTCCAAACAAT	Cebador de inspección inverso para CR36 inespecífico 6
341	CR36-Cell-OT7-F	GGGAATCACAGTAGATGTTGTCA	Cebador de inspección directo para CR36 inespecífico 7
342	CR36-Cell-OT7-R	AGACCAGGAGGTAAGAACATTIG	Cebador de inspección inverso para CR36 inespecífico 7
343	CR36 -Cell-OT8-F	CCACATAGAAAGAGACTTGCAGAA	Cebador de inspección directo para CR36 inespecífico 8
344	CR36-Cell-OT8-R	AGAGATGCCAAAAGAACAGTCAAT	Cebador de inspección inverso para CR36 inespecífico 8
345	CR36-Cell-OT9-F	TGTGCCTTAGGCTATGTAAACTGT	Cebador de inspección directo para CR36 inespecífico 9
346	CR36-Cell-OT9-R	AAACCCTGTAAACCAAAATTACCA	Cebador de inspección inverso para CR36 inespecífico 9
347	CR36-Cell-OT10-F	TAACTGCATCAGAACGCTTGCATA	Cebador de inspección directo para CR36 inespecífico 10
348	CR36-Cell-OT10-R	GGAGACCAAGCTGCTAAAGTCA	Cebador de inspección inverso para CR36 inespecífico 10
349	Cell-CR3-F-anidado	GTGGTGeccggGAGTTGGCTCAAAT TGTTACTCTT	Cebadores de primera ronda de PCR anidada
350	Cell-CR3-R-anidado	GTGGTGeccggGGGAAATGGTCTAG GAGAGTAAAGT	Cebadores de primera ronda de PCR anidada
351	Cell-CR1-F-anidado	GTGGTGeccggGAGAGGTTATGTGGC TTTACCA	Cebadores de primera ronda de PCR anidada
352	Cell-CR1-R-anidado	GTGGTGeccggCTCATTCTCATGCCT GGACA	Cebadores de primera ronda de PCR anidada
353	Cell-CR5-F-anidado	GTGGTGeccggCGGGCTTGGACAGA ACTTAC	Cebadores de primera ronda de PCR anidada
354	Cell-CR5-R-anidado	GTGGTGeccggCTGCGTAGTGCCAA ACAAA	Cebadores de primera ronda de PCR anidada
355	Cell-CR6-F-anidado	GTGGTGeccggTAATTCATTGAAGA GTGGCTGAA	Cebadores de primera ronda de PCR anidada
356	Cell-CR6-R-anidado	GTGGTGeccggAAGCCCTGTGTGGTA GTAGTCAGT	Cebadores de primera ronda de PCR anidada
357	Cell-CR36-F-anidado	GTGGTGeccggCAAGTCAGAACAGTCAC TTGCTTTGT	Cebadores de primera ronda de PCR anidada
358	Cell-CR36-R-anidado	GTGGTGeccggTTTTATGTGCAGGAA TCAGTCTGT	Cebadores de primera ronda de PCR anidada
359	CR3-Cell-OT1-F-anidado	GTGGTGeccggTGTGTGCTCTGTAC ACATCATCT	Cebadores de primera ronda de PCR anidada
360	CR3-Cell-OT1-R-anidado	GTGGTGeccggAGATTCAACCCTCA AAAAGTGAG	Cebadores de primera ronda de PCR anidada
361	CRI-Cell-OT1-F-anidado	GTGGTGeccggTTGGGAACATCAGAG AAAGTATGA	Cebadores de primera ronda de PCR anidada

ES 2 987 399 T3

362	CRI-Cell-OT1-R-anidado	GTGGTGeccggACAAATTACAGTCTC CTGGGAAAG	Cebadores de primera ronda de PCR anidada
363	CR36-Cell-OT3-F-anidado	GTGGTGeccggCACTTCCCAAGTGAG GCAAT	Cebadores de primera ronda de PCR anidada
364	CR36-Cell-OT3-R-anidado	GTGGTGeccggCTATACTTGGGGCTG ACTTGCTAC	Cebadores de primera ronda de PCR anidada
365	CR3-P1/P3-F	GTGGTGeccggTTGGCTCTTAGCTT GTGTTTC	Cebadores de segunda ronda de PCR anidada
366	CR3-P1/P3-R	GTGGTGeccggTGAGACTCCAAAGG CAATC	Cebadores de segunda ronda de PCR anidada
367	CR3-P1/P4-F	GTGGTGeccggTTGGCTCTTAGCTT GTGTTTC	Cebadores de segunda ronda de PCR anidada
368	CR3-P1/P4-R	GTGGTGeccggACTGAGGGGTGATCT TGGTG	Cebadores de segunda ronda de PCR anidada
369	CR3-P2/P3-F	GTGGTGeccggGCAGAGAAAGCCAG TCGGTA	Cebadores de segunda ronda de PCR anidada
370	CR3-P2/P3-R	GTGGTGeccggTGAGACTCCAAAGG CAATC	Cebadores de segunda ronda de PCR anidada
371	CR3-P2/P4-F	GTGGTGeccggGCAGAGAAAGCCAG TCGGTA	Cebadores de segunda ronda de PCR anidada
372	CR3-P2/P4-R	GTGGTGeccggACTGAGGGGTGATCT TGGTG	Cebadores de segunda ronda de PCR anidada
373	CR1-P1/P5-F	GTGGTGeccggCCAGAGTTCTAGGG CAGAG	Cebadores de segunda ronda de PCR anidada
374	CR1-P1/P5-R	GTGGTGeccggAGCTAGTCCCCACAT TCCAC	Cebadores de segunda ronda de PCR anidada
375	CR1-P1/P6-F	GTGGTGeccggCCAGAGTTCTAGGG CAGAG	Cebadores de segunda ronda de PCR anidada
376	CR1-P1/P6-R	GTGGTGeccggGGTGGAGGGAAACT TTAGGC	Cebadores de segunda ronda de PCR anidada
377	CR1-P2/P5-F	GTGGTGeccggCTCATTCTCATGCCT GGACA	Cebadores de segunda ronda de PCR anidada
378	CR1-P2/P5-R	GTGGTGeccggAGCTAGTCCCCACAT TCCAC	Cebadores de segunda ronda de PCR anidada
379	CR1-P2/P6-F	GTGGTGeccggTCTCATGCCTGGACA AGTAACT	Cebadores de segunda ronda de PCR anidada
380	CR1-P2/P6-R	GTGGTGeccggGGTGGAGGGAAACT TTAGGC	Cebadores de segunda ronda de PCR anidada
381	CR5-P3/P5-F	GTGGTGeccggGGCTTGGACAGAACT TACCG	Cebadores de segunda ronda de PCR anidada
382	CR5-P3/P5-R	GTGGTGeccggCACCACTGTCTGCCT AAGGA	Cebadores de segunda ronda de PCR anidada
383	CR5-P4/P6-F	GTGGTGeccggGGCTTGGACAGAACT TACCG	Cebadores de segunda ronda de PCR anidada
384	CR5-P4/P6-R	GTGGTGeccggGGTGGAGGGAAACT TTAGGC	Cebadores de segunda ronda de PCR anidada
385	CR5-P3/P5-F	GTGGTGeccggCGTAGTGCCAAACAA AACAGT	Cebadores de segunda ronda de PCR anidada
386	CR5-P3/P5-R	GTGGTGeccggCACCACTGTCTGCCT AAGGA	Cebadores de segunda ronda de PCR anidada
387	CR5-P4/P6-F	GTGGTGeccggCGTAGTGCCAAACAA AACAGT	Cebadores de segunda ronda de PCR anidada
388	CR5-P4/P6-R	GTGGTGeccggGGTGGAGGGAAACT TTAGGC	Cebadores de segunda ronda de PCR anidada
389	CR6-P1/P5-F	GTGGTGeccggGCGAGGGCCTACTTG ATATG	Cebadores de segunda ronda de PCR anidada

390	CR6-P1/P5-R	GTGGTGeccggCTTCCAAGTGAGGC AATGC	Cebadores de segunda ronda de PCR anidada
391	CR6-P1/P6-F	GTGGTGeccggACGTTTGTGCTGCT GTAACA	Cebadores de segunda ronda de PCR anidada
392	CR6-P1/P6-R	GTGGTGeccggCTGCAGGCACATTCT CTTCC	Cebadores de segunda ronda de PCR anidada
393	CR6-P2/P5-F	GTGGTGeccggGCCCTGTGTGGTAGT AGTCA	Cebadores de segunda ronda de PCR anidada
394	CR6-P2/P5-R	GTGGTGeccggCTTCCAAGTGAGGC AATGC	Cebadores de segunda ronda de PCR anidada
395	CR6-P2/P6-F	GTGGTGeccggCAGTATTAAGGGGTG GGAGCT	Cebadores de segunda ronda de PCR anidada
396	CR6-P2/P6-R	GTGGTGeccggTCTCTTCACACA GCTGA	Cebadores de segunda ronda de PCR anidada
397	CR36-P3/P5-F	GTGGTGeccggGGAGCTGGAGGGA AGAGAA	Cebadores de segunda ronda de PCR anidada
398	CR36-P3/P5-R	GTGGTGeccggCTTCCAAGTGAGGC AATGC	Cebadores de segunda ronda de PCR anidada
399	CR36-P4/P6-F	GTGGTGeccggATGGATGGGAAGA CACTGG	Cebadores de segunda ronda de PCR anidada
400	CR36-P4/P6-R	GTGGTGeccggCTGCAGGCACATTCT CTTCC	Cebadores de segunda ronda de PCR anidada
401	CR36-P3/P5-F	GTGGTGeccggGGATGAAACAGGGC AGGAAC	Cebadores de segunda ronda de PCR anidada
402	CR36-P3/P5-R	GTGGTGeccggTTCCAAGTGAGGCA ATGC	Cebadores de segunda ronda de PCR anidada
403	CR36-P4/P6-F	GTGGTGeccggTTTGAGAGCCATGA TGAGG	Cebadores de segunda ronda de PCR anidada
404	CR36-P4/P6-R	GTGGTGeccggCGACAGCCAAACAA GCCG	Cebadores de segunda ronda de PCR anidada

Análisis de transferencia Western. Para evaluar la expresión de la proteína distrofina, los mioblastos inmortalizados se diferenciaron en miofibra reemplazando el medio de crecimiento con DMEM suplementado con insulina-transferrina-selenio al 1 % (Invitrogen) y antibiótico/antimicótico al 1 % (Invitrogen) durante 4-7 días, tal como 6 o

5 7 días. Los fibroblastos se transdiferenciaron en mioblastos induciendo la sobreexpresión de MyoD e incubando las células en DMEM suplementado con insulina-transferrina-selenio al 1 % (Invitrogen), antibiótico/antimicótico al 1 % (Invitrogen) y doxiciclina 3 µg/ml durante 15 días. La expresión de distrofina se evaluó 3 días después de transfectar células HEK293T. Las células se trisnizaron, se recogieron y se lisaron en tampón RIPA (Sigma) suplementado con un cóctel inhibidor de proteasa (Sigma) y la cantidad total de proteína se cuantificó usando el ensayo de ácido bicinconíco de acuerdo con las instrucciones del fabricante (Pierce). Las muestras se mezclaron después con tampón de carga NuPAGE (Invitrogen) y β-mercaptoetanol al 5 % y se calentaron a 85 °C durante 10 minutos. Se separaron veinticinco microgramos de proteína en geles NuPAGE Bis-Tris al 4-12 % (Invitrogen) con tampón MES (Invitrogen). Las proteínas se transfirieron a membranas de nitrocelulosa durante 1-2 h en un tampón de transferencia que contenía metanol al 10-20 %, tal como metanol al 10 % y SDS al 0,01 %. La transferencia se bloqueó después durante 1 h con 10 TBST de leche al 5 % a temperatura ambiente. Las transferencias se probaron con los siguientes anticuerpos primarios: MANDYS8 para detectar distrofina (1:100, Sigma D8168) y anti-GAPDH de conejo (1:5000, Cell Signaling 2118S). Las transferencias se incubaron después con anticuerpos secundarios de ratón o conejo conjugados con peroxidasa de rábano picante (Santa Cruz) y se visualizaron usando el sistema quimioluminiscente ChemiDoc (BioRad) y el sustrato Western-C ECL (BioRad).

20 **Trasplante a ratones inmunodeficientes.** Todos los experimentos con animales se realizaron según protocolos aprobados por el Comité Institucional de Cuidado y Uso de Animales de Duke. Las células se trisnizaron, se recogieron y se lavaron en solución salina equilibrada de Hank 1X (HBSS, Sigma). Se sedimentaron y se resuspendieron dos millones de células en cinco µl de HBSS IX (Sigma) suplementado con cardiotoxina (Sigma n.º C9759) inmediatamente antes de la inyección. Estas células se trasplantaron en el músculo tibial anterior (TA) de la extremidad trasera de ratones NOD.SCID.gamma (NSG) (Duke CCIF Breeding Core) mediante inyección intramuscular. Cuatro semanas después de la inyección, los ratones se sacrificaron y se recogieron los músculos TA.

30 **Tinción por inmunofluorescencia.** Los músculos TA recogidos se incubaron en glicerol al 30 % durante la noche a 4 °C antes de montarlos y congelarlos en un compuesto de temperatura de corte óptima. Se obtuvieron secciones seriadas de 10 micrómetros mediante criosección del tejido muscular embebido a -20 °C. Las criosecciones se lavaron

después en PBS para eliminar el compuesto OCT y posteriormente se bloquearon durante 30-60 minutos a temperatura ambiente en PBS que contenía un 10 % de suero bovino fetal inactivado por calor para la detección de espectrina o suero bovino fetal al 5 % inactivado por calor para la detección de distrofina. Las criosecciones se incubaron durante la noche a 4 °C con los siguientes anticuerpos primarios que son específicos únicamente de epitopos humanos: antiespectrina (1:20, Leica NCL-SPEC1) o antidistrofina (1:2, Leica NCL-DYS3). Después de la tinción primaria, la expresión de espectrina o distrofina se detectó usando un kit de detección de amplificación de señal de inmunofluorescencia basado en tiramida (Life Technologies, TSA Kit n.º 22, catálogo n.º T-20932,). Brevemente, las criosecciones se incubaron con cabra anti-biotina-XX de ratón 1:200 secundaria (Life Technologies N.º B2763) en tampón de bloqueo durante 1 h a temperatura ambiente. La señal se amplificó después usando conjugados de estreptavidina-HRP (1:100, del kit TSA) en tampón de bloqueo durante 1 h a temperatura ambiente. Finalmente, las criosecciones se incubaron con conjugados de tiramida-AlexaFluor488 (1:100, kit TSA) en el tampón de amplificación proporcionado por el fabricante durante 10 minutos a temperatura ambiente. Las criosecciones teñidas se montaron después en ProLong AntiFade (Life Technologies N.º P36934) y se visualizaron con microscopía de fluorescencia convencional.

15 **Ensayo de citotoxicidad.** Para evaluar cuantitativamente la posible citotoxicidad asociada a ARNgm o nucleasa SpCas9, las células HEK293T se transfecaron con 10 ng de un indicador GFP y 100 ng del vector de expresión SpCas9 y 100 ng del vector de expresión ARNgm usando Lipofectamine 2000 de acuerdo con las instrucciones del fabricante (Invitrogen). El porcentaje de células positivas para GFP se evaluó a los 2 y 5 días mediante citometría de flujo. La tasa de supervivencia se calculó como la disminución de células positivas para GFP de los días 2 a 5 y se normalizó para células transfectadas con un vector de expresión de nucleasa vacío como se describe (Cornu *et al.*, Meth Mol Biol 649:237-245 (2010)).

Ejemplo 4

CRISPR dirigidos al gen de la distrofina: resultados

El sistema basado en CRISPR/Cas9 se diseñó para dirigirse al gen de la distrofina. Se eligieron varios ARNg para dirigirse a diferentes regiones del gen de la distrofina humana y de ratón basándose en NNNNN NNNNN NNNNN NNNNN NNNNN NGG y GNNNN NNNNN NNNNN NNNNN NGG (véanse las Tablas 6, 7 y 8).

Tabla 6

Nombr e (SEQ ID NO)	Especie	Gen	Diana	Cadena	19 pb para químérico (añadir G en el extremo 5')	PAM
DCR1 (65)	Ser humano	DMD	Intrón 50	+	attgccttgcattcccta	GGG
DCR2 (66)	Ser humano	DMD	Intrón 50	-	tgttagatgttcagccata	TGG
DCR3 (67)	Ser humano	DMD	Exón 51-55'	+	cctactcagactgttaactc	TGG
DCR4 (68)	Ser humano	DMD	Exón 51-53'	+	ttggacagaaccttaccgac	TGG
DCR5 (69)	Ser humano	DMD	Intrón 51	-	cagtgcctaagaactggt	GGG
DCR6 (70)	Ser humano	DMD	Intrón 44	-	GGGCTCCACCCCTCACGAGT	GGG
DCR7 (71)	Ser humano	DMD	Intrón 55	+	TTTGCTTCGCTATAAAACG	AGG
DCR8 (72)	Ser humano	DMD	Exón 41	+	TCTGAGGATGGGCCGCAA	TGG
DCR9 (73)	Ser humano	DMD	Exón 44	-	GATCTGTCAAATCGCCTGC	AGG
DCR10 (74)	Ser humano	DMD	Exón 45	+	CCAGGATGGCATTGGCAG	CGG
DCR11 (75)	Ser humano	DMD	Exón 45	+	CTGAATCTGCGGTGGCAGG	AGG
DCR12 (76)	Ser humano	DMD	Exón 46	-	TTCTTTGTTCTTAGCc	TGG
DCR13 (77)	Ser humano	DMD	Exón 46	+	GAAAAGCTTGAGCAAGTCA	AGG

ES 2 987 399 T3

DCR14 (78)	Ser humano	DMD	Exón 47	+	GAAGAGTTGCCCTGCGCC	AGG
DCR15 (79)	Ser humano	DMD	Exón 47	+	ACAAATCTCCAGTGGATAA	AGG
DCR16 (80)	Ser humano	DMD	Exón 48	-	TGTTTCTCAGGTAAAGCTC	TGG
DCR17 (81)	Ser humano	DMD	Exón 48	+	GAAGGACCATTGACGTTa	AGG
DCR18 (82)	Ser humano	DMD	Exón 49	-	AACTGCTATTCAGTTTCc	TGG
DCR19 (83)	Ser humano	DMD	Exón 49	+	CCAGCCACTCAGCCAGTGA	AGG
DCR20 (84)	Ser humano	DMD	Exón 50	+	gtatgcctttctgttaaaag	AGG
DCR21 (85)	Ser humano	DMD	Exón 50	+	CTCCTGGACTGACCACTAT	TGG
DCR22 (86)	Ser humano	DMD	Exón 52	+	GAACAGAGGCGTCCCCAGT	TGG
DCR23 (87)	Ser humano	DMD	Exón 52	+	GAGGCTAGAACAAATCATTA	CGG
DCR24 (88)	Ser humano	DMD	Exón 53	+	ACAAGAACACCTTCAGAAC	CGG
DCR25 (89)	Ser humano	DMD	Exón 53	-	GGTTTCTGTGATTTCTTT	TGG
DCR26 (90)	Ser humano	DMD	Exón 54	+	GGCCAAAGACCTCCGCCAG	TGG
DCR27 (91)	Ser humano	DMD	Exón 54	+	TTGGAGAAGCATTCAAAA	AGG
DCR28 (92)	Ser humano	DMD	Exón 55	-	TCGCTCACTCACCCtgc当地	AGG
DCR29 (93)	Ser humano	DMD	Exón 55	+	AAAAGAGCTGATGAAACAA	TGG
DCR30 (94)	Ser humano	DMD	5'UTR/Exón 1	+	TAcACTTTCaAAATGCTT	TGG
DCR31 (95)	Ser humano	DMD	Exón 51	+	gagatgatcatcaaggcaga	AGG
DCR32 (96)	Ratón	DMD	mutación mdx	+	ctttggaaagagcaaTaaaa	TGG
DCR33 (97)	Ser humano	DMD	Intrón 44	-	CACAAAAGTCAAATCGGAA	TGG
DCR34 (98)	Ser humano	DMD	Intrón 44	-	ATTTCAATATAAGATTGG	AGG
DCR35 (99)	Ser humano	DMD	Intrón 55	-	CTTAAGCAATCCGAACTC	TGG
DCR36 (100)	Ser humano	DMD	Intrón 55	-	CCTTCTTTATCCCTATCG	AGG
DCR40 (104)	Ratón	DMD	Exón 23	-	aggccaaacctcggttac	NNGRR
DCR41 (105)	Ratón	DMD	Exón 23	+	TTCGAAAATTCAAGgttag	NNGRR
DCR42 (106)	Ratón	DMD	Exón 23	+	gcagaacaggagataacag	NNGRRT
DCR43 (107)	Ratón	ACVR2 B	Exón 1	+	gcggccctcgcccttcct	ggggat
DCR48 (108)	Ser humano	DMD	Intrón 45	-	TAGTGATCGTGGATACGAG	AGG

ES 2 987 399 T3

DCR49 (109)	Ser humano	DMD	Intrón 45	-	TACAGCCCTCGGTATAT	TGG
DCR50 (110)	Ser humano	DMD	Intrón 52	-	GGAAGGAATTAAAGCCCCGAA	TGG
DCR51 (111)	Ser humano	DMD	Intrón 53	-	GGAACAGCTTCGTAGTTG	AGG
DCR52 (112)	Ser humano	DMD	Intrón 54	+	ATAAAGTCCAGTGTCGATC	AGG
DCR53 (113)			Intrón 54	+	AAAACCAGAGCTTCGGTCA	AGG
DCR54 (114)	Ratón	Rosa26	Región ZFN	+	GAGTCTTCTGGGCAGGCTTAA AGGCTAAC	TGG
DCR55 (115)	Ratón	Rosa26	ARNm	-	TCGGGTGAGCATGTCTTAAT CTACUTCGA	TGG
DCR49 (116)	Ser humano	DMD	Ex 51	-	gtgtcaccagagaacagt	ctgagt
DCR50 (117)	Ser humano	DMD	Ex 51	+	tgcgtcatcaaggcagaagg	atgag
DCR60 (118)	Ratón	DMD	Exón 23	+	AACTTCGAAAATTCAGgt	agccgagg
DCR61 (119)	Ratón	DMD	Intrón 22	+	gaaactcatcaaatacgct	gttagtgt
DCR62 (120)	Ratón	DMD	Intrón 22	-	tcatttacactaacacgcat	atttgtatg
DCR63 (121)	Ratón	DMD	Intrón 22	+	gaatgaaactcatcaaata	gcgtgtta
DCR64 (122)	Ratón	DMD	Intrón 23	-	tcatcaataatcttgaagg	cctgggt
DCR65 (123)	Ratón	DMD	Intrón 23	-	tgttttcataggaaaaatag	gcaaggtg
DCR66 (124)	Ratón	DMD	Intrón 23	+	aattggaaaatgtgtgg	aacagata
DCR67 (125)	Ser humano	DMD	Exón 51	+	atgtatcatcaaggcagaagg	atgagaaa
DCR68 (126)	Ser humano	DMD	Exón 51	+	agatgtatcatcaaggcagaag	gtatgaga
DCR69 (127)	Ser humano	DMD	Exón 51	-	catttttcatacccttct	gtttgtatg
DCR70 (128)	Ser humano	DMD	Exón 51	+	tcctactcagactgtactc	tggtgaca
DCR71 (129)	Ser humano	DMD	Exón 51	-	acagggttgtcaccagagt	aacagtct
DCR72 (130)	Ser humano	DMD	Exón 51	-	ttatcatttttctcatacc	ttctgttt
DCR73 (131)	Ser humano	DMD	Intrón 51	-	ttgcctaagaactgggg	aatggct
DCR74 (132)	Ser humano	DMD	Intrón 51	-	aaacagttgectaagaactg	gtgggaaa
DCR75 (133)	Ser humano	DMD	Intrón 51	+	tttcccaccaggtttaggc	aactgttt
DCR76 (134)	Ser humano	DMD	Intrón 50	+	tggcttgattccctaggg	tccagett
DCR77 (135)	Ser humano	DMD	Intrón 50	-	tagggaaatcaaaggcaal	aaacgttc
DCR78 (136)	Ser humano	DMD	Intrón 50	-	gacccttagggaaatcaaagc	caatgaaa

DCR79 (137)	Ser humano	DMD	Intrón 44	-	TGAGGGCTCCACCCACGA	GTGGGTTT
DCR80 (138)	Ser humano	DMD	Intrón 44	-	AAGGATTGAGGGCTCCACCC	TCACGAGT
DCR81 (139)	Ser humano	DMD	Intrón 44	-	GCTCCACCCACGAGTGCG	TTTGGTTC
DCR82 (140)	Ser humano	DMD	Intrón 55	-	TATCCCCTATCGAGGAAACC	ACGAGTTT
DCR83 (141)	Ser humano	DMD	Intrón 55	+	GATAAAGAAGGCCTATTCA	TAGAGTTG
DCR84 (142)	Ser humano	DMD	Intrón 55	-	AGGCCTTCTTATCCCCTAT	CGAGGAAA
DCR85 (143)	Ser humano	DMD	Intrón 44	-	TGAGGGCTCCACCCACGA	GTGGGT
DCR86 (144)	Ser humano	DMD	Intrón 55	+	GATAAAGAAGGCCTATTCA	TAGAGT

Tabla 7

Nombre	Notas	% de Mod
DCR1	Eliminar el exón 51	6,6
DCR2	Eliminar el exón 51	10,3
DCR3	Desplazamiento de marco	13
DCR4	Eliminar el exón 51	11,9
DCR5	Eliminar el exón 51	12,4
DCR6	Lo más cerca posible del exón 44 en el intrón 44 (en caso de eliminaciones del paciente)	16,1
DCR7	Lo más cerca posible del exón 56 en el intrón 55 (en caso de eliminaciones del paciente)	6,8
DCR8	Se puede corregir la eliminación del exón 42-43 (-1/+2) únicamente, (-2/+1) no puede corregirse con esto	17,3
DCR9	Saltar el exón 44 (5')	14,4
DCR10	Desplazamiento de marco	14,9
DCR11	Corregir en dirección 3' del exón 45	<1
DCR12	Aceptor de corte y empalme/desplazamiento de marco 5'	<1
DCR13	Corregir en dirección 3' del exón 46	16,9
DCR14	Desplazamiento de marco	17,2
DCR15	Corregir en dirección 3' del exón 47	15,4
DCR16	Desplazamiento de marco	11,5
DCR17	Corregir en dirección 3' del exón 48	<1
DCR18	Aceptor de corte y empalme/desplazamiento de marco 5'	1,8
DCR19	Corregir en dirección 3' del exón 49	33,7
DCR20	Aceptor de corte y empalme 5'	14,9
DCR21	Corregir en dirección 3' del exón 50	24,1
DCR22	Desplazamiento de marco	25,9
DCR23	Corregir en dirección 3' del exón 52	25,2
DCR24	Desplazamiento de marco (solo puede corregirse marco + 1)	24,8
DCR25	Corregir en dirección 3' del exón 53	2,6
DCR26	Desplazamiento de marco	24,5
DCR27	Corregir en dirección 3' del exón 54	13,4
DCR28	Aceptor de corte y empalme 5'	21,6
DCR29	Corregir en dirección 3' del exón 55	19,2
DCR30	Integrar minidys en el exón 1	no probado
DCR31	Corregir en dirección 3' del exón 51	18,9

DCR32	Eliminar el codón de parada	no probado
DCR33	Alternativa a CR6	1,3
DCR34	Alternativa a CR6	13,2
DCR35	Alternativa a CR7	22,5
DCR36	Alternativa a CR7	26,4
DCR40	Interrumpir el donante de corte y empalme 5' del exón 23 (corregir mutación mdx)	
DCR41	Interrumpir el donante de corte y empalme 5' del exón 23 (corregir mutación mdx)	
DCR42	Eliminar la mutación mdx4cv del exón 53	
DCR43	Interrumpir el receptor de miostatina	

Tabla 8

Nombre	Cas9		Notas	Cas9 usada
DCR49	<i>S. aureus</i>		Desplazamiento de marco en el exón 51	SaCas9 (de Zhang pX441) (NNGRRT PAM)
DCR50	<i>S. aureus</i>		Interrumpir el extremo 5' del exón 51	SaCas9 (de Zhang pX441) (NNGRR PAM)
DCR60	<i>N. meningitidis</i>	NNNNGANN	Dirigirse al donante de corte y empalme 3' del exón 23 para evitar la mutación mdx	NmCas9 (NNNNGANN PAM)
DCR61	<i>N. meningitidis</i>	NNNNGNNT	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGNNT PAM)
DCR62	<i>N. meningitidis</i>	NNNNGANN	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGANN PAM)
DCR63	<i>N. meningitidis</i>	NNNNGTTN	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGTTN PAM)
DCR64	<i>N. meningitidis</i>	NNNNGNNT	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGNNT PAM)
DCR65	<i>N. meningitidis</i>	NNNNGTTN	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGTTN PAM)
DCR66	<i>N. meningitidis</i>	NNNNGANN	Eliminar el exón 23 y evitar la mutación mdx	NmCas9 (NNNNGANN PAM)
DCR67	<i>N. meningitidis</i>	NNNNGANN	Dirigirse al donante de corte y empalme 3' del exón 51 para omitir el exón	NmCas9 (NNNNGANN PAM)
DCR68	<i>N. meningitidis</i>	NNNNGANN	Dirigirse al donante de corte y empalme 3' del exón 51 para omitir el exón	NmCas9 (NNNNGANN PAM)
DCR69	<i>N. meningitidis</i>	NNNNGANN	Dirigirse al donante de corte y empalme 3' del exón 51 para omitir el exón	NmCas9 (NNNNGANN PAM)
DCR70	<i>N. meningitidis</i>	NNNNGANN	Desplazamiento de marco en el exón 51	NmCas9 (NNNNGANN PAM)
DCR71	<i>N. meningitidis</i>	NNNNGNNT	Desplazamiento de marco en el exón 51	NmCas9 (NNNNGNNT PAM)
DCR72	<i>N. meningitidis</i>	NNNNGNNT	Dirigirse al donante de corte y empalme 3' del exón 51 para omitir el exón	NmCas9 (NNNNGNNT PAM)
DCR73	<i>N. meningitidis</i>	NNNNGNNT	Eliminar el exón 51 (unirse lo más cerca posible de DCR5)	NmCas9 (NNNNGNNT PAM)
DCR74	<i>N. meningitidis</i>	NNNNGANN	Eliminar el exón 51 (unirse lo más cerca posible de DCR5)	NmCas9 (NNNNGANN PAM)
DCR75	<i>N. meningitidis</i>	NNNNGTTN	Eliminar el exón 51 (unirse lo más cerca posible de DCR5)	NmCas9 (NNNNGTTN PAM)
DCR76	<i>N. meningitidis</i>	NNNNGNNT	Eliminar el exón 51 (unirse lo más cerca posible de DCR1/2)	NmCas9 (NNNNGNNT PAM)

DCR77	<i>N. meningitidis</i>	NNNNNGTTN	Eliminar el exón 51 (unirse lo más cerca posible de DCR1/2)	NmCas9 (NNNNNGTTN PAM)
DCR78	<i>N. meningitidis</i>	NNNNNGANN	Eliminar el exón 51 (unirse lo más cerca posible de DCR1/2)	NmCas9 (NNNNNGANN PAM)
DCR79	<i>N. meningitidis</i>	NNNNNGNNT	Eliminar los exones 45-55: se superpone a NNNNGTTN PAM, unirse lo más cerca posible de DCR6	NmCas9 (NNNNGNNT PAM)
DCR80	<i>N. meningitidis</i>	NNNNNGANN	Eliminar los exones 45-55, unirse lo más cerca posible de DCR6	NmCas9 (NNNNNGANN PAM)
DCR81	<i>N. meningitidis</i>	NNNNNGTTN	Eliminar los exones 45-55, unirse lo más cerca posible de DCR6	NmCas9 (NNNNNGTTN PAM)
DCR82	<i>N. meningitidis</i>	NNNNGNNT	Eliminar los exones 45-55, unirse lo más cerca posible de DCR36: se superpone a NNNNGTTN PAM	NmCas9 (NNNNGNNT PAM)
DCR83	<i>N. meningitidis</i>	NNNNNGTTN	Eliminar los exones 45-55, unirse lo más cerca posible de DCR36	NmCas9 (NNNNNGTTN PAM)
DCR84	<i>N. meningitidis</i>	NNNNNGANN	Eliminar los exones 45-55, unirse lo más cerca posible de DCR36	NmCas9 (NNNNNGANN PAM)
DCR85	<i>S. aureus</i>	NNGRRT	Eliminar los exones 45-55, unirse lo más cerca posible de DCR6	SaCas9 (de Zhang pX441) (NNGRRT PAM)
DCR86	<i>S. aureus</i>	NNGRRT	Eliminar los exones 45-55, unirse lo más cerca posible de DCR36	SaCas9 (de Zhang pX441) (NNGRRT PAM)

En particular, se cotransfектaron 400 ng de Cas9 en células HEK 293T con 400 ng de vector vacío o ARNg que se dirige a la región que abarca el exón 51, es decir, CR1, CR2, CR3, CR4 y CR5 (véase la Fig. 11(b)). El ADN genómico se recogió 2 días después de la transfección y se analizó usando el ensayo de inspección (véanse las Fig. 11(a) y 11(c)).

5

El sistema basado en CRISPR/Cas9 se usó en células DMD 8036 (del48-50) para determinar si el sistema podía reparar un gen de distrofina mutante. Se cotransfектaron 5 µg de Cas9 en células DMD 8036 (del48-50) con 7,5 µg de vector vacío o ARNg. En particular, se usaron 7,5 µg de CR1 ("DCR1"), 7,5 µg de CR5 ("DCR5"), 15 µg de CR3 ("DCR3") o 7,5 µg de una combinación de CR1 y CR5 (DCR1+DCR5). El ADN genómico se recogió 3 días después de la transfección y se analizó mediante el ensayo de inspección (Fig. 12) o análisis de PCR en todo el locus (Fig. 13). Este locus se amplificó mediante PCR usando cebadores que flanquean la región que contiene las dianas genómicas para CR1 y CR5 (el cebador directo: 5'-gagaggatgtggcttacca (SEQ ID NO:457), el cebador inverso: 5'-ctgcgttagtgccaaaacaaa (SEQ ID NO:458)), dando como resultado una banda de 1447 pb para el locus de tipo silvestre o un tamaño esperado de aproximadamente 630 pb para el locus eliminado. Después de 7 días de diferenciación, la transferencia Western de las células tratadas muestra la expresión de la proteína distrofina (véase la Fig. 14).

10

15

Ejemplo 5

20 Direcciónamiento de CRISPR/Cas9 a mutaciones de puntos críticos en el gen de la distrofina humana

Para utilizar la plataforma de edición génica CRISPR/Cas9 para corregir una amplia gama de mutaciones de distrofina, se crearon docenas de ARNg dirigidos a la región de mutación del punto de acceso entre los exones 45-55 (Fig. 16). Se usó el sistema de *S. pyogenes* que utiliza una nucleasa SpCas9 optimizado por codones humanos y un vector de expresión de ARN guía monocatenario (ARNgm) químérico para guiar la edición génica específica de sitio eficiente. Similar al Ejemplo 4 dirigido al exón 51 con TALEN, se seleccionaron protoespaciadores para dirigirse a los extremos 5' y 3' de los exones 45 a 55 que cumplen con el requisito PAM 5'-NRG-3' de SpCas9. Pueden generarse pequeñas inserciones o eliminaciones creadas por la reparación del ADN basada en NHEJ dentro de estos exones pueden generar mutaciones de cambio de marco específicas que abordan diversas mutaciones de distrofina que rodean cada exón (Fig. 16A-16B). Por ejemplo, CR3 se diseñó para corregir mutaciones o eliminaciones de distrofina que rodean el exón 51 mediante la introducción de pequeñas inserciones o eliminaciones en el extremo 5' del exón 51 para restaurar el marco de lectura de distrofina en dirección 3' (Fig. 16B). Adicionalmente, los ARNg se diseñaron para emplear la capacidad multiplex del sistema CRISPR/Cas9 y eliminar específicamente exones individuales o una serie

de exones para restaurar el marco de lectura de distrofina, similar a los métodos de omisión de exones basados en oligonucleótidos. Para este fin, los ARN_{gm} se dirigieron a las regiones intrónicas que rodean el exón 51 (Fig. 16C) o los exones 45-55 (Fig. 16D). Estos ARN_{gm} se dirigieron intencionalmente a los sitios más cercanos al exón en dirección 3' o ascendente que se pretendía incluir en la transcripción resultante para minimizar la probabilidad de que la eliminación de fondo del paciente incluyera los sitios diana de ARN_{gm} intrónicos.

Ejemplo 6

Detección de ARN_{gm} dirigidos al gen de la distrofina en células humanas

Se evaluó la frecuencia de edición génica en la línea celular humana HEK293T para determinar rápidamente diferentes eficiencias de direccionamiento de ARN_{gm}. Las HEK293T se transfecaron con construcciones que codifican SpCas9 con codones humanos optimizados y el ARN_{gm} indicado. Cada ARN_{gm} se diseñó para modificar el gen de la distrofina como se indica. La frecuencia de modificación génica en el día 3 o 10 después de la transfección se determinó mediante el ensayo de inspección. La proporción de la señal de inspección medida en los días 3 y 10 se calculó para cuantificar la estabilidad de las frecuencias de edición génica para cada ARN_{gm} en células humanas. Según lo cuantificado por el ensayo de inspección 3 días después de la transfección, 29/32 (~90 %) de los ARN_{gm} analizados pudieron mediar una modificación génica altamente eficiente en el locus deseado (Tabla 9, Fig. 17). Las frecuencias de edición génica fueron estables para casi todos los ARN_{gm} (cambio de señal <25 % del día 3 al día 10, Tabla 9, Fig. 18), lo que indica que la edición génica mediada por cada ARN_{gm} individual fue bien tolerada. Una excepción notable es CR33, que no tuvo actividad detectable el día 10, aunque la actividad puede estar por debajo de la sensibilidad del ensayo de inspección (est. ~1 %).

Tabla 9 Actividad medida de ARN_{gm} en células humanas

Diana	N.º de ARN _{gm}	% de alelos modificados el día 3	% de alelos modificados el día 10	% de cambio día 10/día 3
Eliminación múltiplex del exón 51				
Int 50	CR1	6,6	9,3	41,8
Int 50	CR2	10,3	14,0	36,2
Ex 51	CR4	11,9	14,4	21,3
Int 51	CR5	12,4	13,3	7,8
Eliminación múltiplex de los exones 45-55				
Int 44	CR6	16,1	16,9	4,3
Int 44	CR33	1,3	<1	n.d.
Int 44	CR34	13,2	11,0	-16,6
Int 55	CR7	6,8	7,1	5,3
Int 55	CR35	22,5	20,9	-7,1
Int 55	CR36	26,4	24,7	-6,4
Desplazamientos de marco direccionalizados				
Ex 45	CR10	14,9	16,3	9,3
Ex 45	CR11	<1	<1	n.d.
Ex 46	CR12	<1	<1	n.d.
Ex 46	CR13	16,9	18,4	9,2
Ex 47	CR14	17,2	17,6	2,9
Ex 47	CR15	15,4	15,3	-0,9
Ex 48	CR16	11,5	10,9	-5,0
Ex 48	CR17	<1	<1	n.d.
Ex 49	CR18	1,8	2,2	20,1
Ex 49	CR19	33,7	38,4	13,9
Ex 50	CR20	14,9	13,7	-7,6
Ex 50	CR21	24,1	20,8	-13,5
Ex 51	CR3	13,0	16,7	28,0
Ex 51	CR31	18,9	16,9	-10,2
Ex 52	CR22	25,9	20,3	-21,6

Ex 52	CR23	25,2	24,0	-4,8
Ex 53	CR24	24,8	23,6	-4,6
Ex 53	CR25	2,6	2,9	9,5
Ex 54	CR26	24,5	22,0	-10,1
Ex 54	CR27	13,4	12,6	-5,9
Ex 55	CR28	21,6	19,8	-8,4
Ex 55	CR29	19,2	19,6	2,2

Ejemplo 7**Enriquecimiento de células editadas génicamente mediante un sistema indicador basado en fluorescencia**

5 Se seleccionaron ARNgm para corregir mutaciones específicas en líneas celulares de mioblastos de pacientes con DMD. Después de la transfección en mioblastos DMD, se observó una actividad de modificación génica inesperadamente baja o indetectable medida por el ensayo de inspección (Fig. 19C, población a granel). Se usó citometría de flujo para seleccionar células transfectadas que coexpresan GFP a través de un péptido de salto 10 ribosómico 2A unido a la proteína SpCas9 (Fig. 19A). La adición de este indicador fluorescente al vector de expresión SpCas9 no pareció afectar significativamente la actividad de edición génica en células HEK293T (Fig. 19B). Un bajo porcentaje de mioblastos transfectados (~0,5-2 %) expresaron el indicador fluorescente 3 días después de la electroporación, a pesar de las altas eficiencias de transfección de los plásmidos de expresión de GFP de control 15 (normalmente >70 %, Fig. 19D, pmaxGFP). Dados los altos niveles de actividad CRISPR/Cas9 en la línea HEK293T fácilmente transfectada, la expresión transgénica ineficiente después de la electroporación de construcciones SpCas9-T2A-GFP y ARNgm en las células DMD puede explicar las bajas eficiencias de edición génica observadas en células 20 no clasificadas. Después de clasificar los mioblastos DMD positivos para GFP, se observó un aumento sustancial en la actividad detectable en la mayoría de los loci objetivo de ARNgm (Fig. 19C). Por lo tanto, todos los experimentos posteriores usaron células clasificadas para la expresión de SpCas9 mediante la expresión de este indicador fluorescente.

Ejemplo 8**Restauración de la expresión de distrofina mediante desplazamientos de marco específicos**

25 Pueden usarse pequeñas inserciones y eliminaciones creadas por la reparación del ADN de NHEJ para crear desplazamientos de marco específicos para corregir marcos de lectura aberrantes. Un ARNg, CR3, se diseñó para restaurar el marco de lectura de distrofina mediante la introducción de pequeñas inserciones y eliminaciones dentro del exón 51 (Figs. 16B, 20A). Los tipos de inserciones y eliminaciones generadas por CRISPR/Cas9 en este locus se 30 evaluaron mediante secuenciación de Sanger de alelos del ADN genómico de células HFK293T cotransfectadas con plásmidos de expresión para SpCas9 y el ARNg de CR3 (Fig. 20B). Notablemente, las inserciones y eliminaciones dieron como resultado la conversión a los tres marcos de lectura (Figs. 20B, 20C). Para demostrar la corrección 35 genética en una línea celular de paciente relevante, los plásmidos de expresión para SpCas9 y el ARNg de CR3 se electroporaron en una línea de mioblastos DMD con una eliminación de los exones 48-50 que puede corregirse mediante la creación de cambios de marco en el exón 51. Las células tratadas se clasificaron, se verificó que tenía actividad de modificación génica mediante el ensayo de inspección (CR3, población clasificada Fig. 19C) y se 40 diferenciaron en miotubos para probar la expresión de distrofina restaurada. Se observó expresión de la proteína distrofina concordantemente con la actividad nucleasa detectable (Fig. 20D). El sistema CRISPR/Cas9 de *S. pyogenes* presenta un método poderoso para generar rápidamente desplazamientos de marco específicos para abordar una diversidad de mutaciones de pacientes y restaurar la expresión del gen de la distrofina humana.

Ejemplo 9**La edición genética multiplex CRISPR/Cas9 media la eliminación genómica del exón 51 y rescata la expresión de la proteína distrofina**

La capacidad de multiplexación del sistema CRISPR/Cas9 presenta un método novedoso para generar de manera 45 eficiente eliminaciones genómicas de exones específicos para la corrección génica específica. Los mioblastos de pacientes con DMD con eliminaciones de fondo corregibles mediante la omisión del exón 51 se trataron con dos combinaciones de ARNg que flanquean el exón 51 (CR1/CR5 o CR2/CR5) y se clasificaron para Enriquecer las 50 células editadas con genes como en la Fig. 19. Según lo detectado por PCR de punto final del ADN genómico de estas células tratadas, las eliminaciones genómicas esperadas solo estuvieron presentes cuando ambos ARNg se electroporaron en las células con SpCas9 (Fig. 21A). La secuenciación de Sanger confirmó la unión esperada de los 55 segmentos cromosómicos distales (Fig. 21B) para ambas eliminaciones. Después de diferenciar los mioblastos clasificados, se detectó una eliminación del exón 51 del transcripto de ARNm solo en las células tratadas con ambos ARNg (Fig. 21C). Finalmente, se detectó la expresión restaurada de la proteína distrofina en las células tratadas de

manera concomitante con las eliminaciones observadas a nivel del genoma y del ARNm del exón 51 (Fig. 21D).

Ejemplo 10

5 Rescate de distrofina mediante una eliminación genómica grande de múltiples exones

Aunque abordar mutaciones específicas de cada paciente es un uso poderoso del sistema CRISPR/Cas9, sería ventajoso desarrollar un método único que pueda abordar una gran cantidad de eliminaciones comunes de pacientes. Por ejemplo, una estrategia prometedora es excluir toda la región de los exones 45-55 como método para corregir hasta el 62 % de las eliminaciones conocidas de los pacientes. Se probó la edición génica multiplex basada en CRISPR/Cas9 para determinar si puede generar una eliminación eficiente del locus del exón 45-55 en células humanas. Después de la transfección en células HEK293T, la eliminación esperada de ~336.000 pb se detectó mediante PCR del ADN genómico (Fig. 22A). De forma similar, esta eliminación se detectó mediante PCR del ADN genómico de células de pacientes con DMD tratadas con SpCas9/ARNgm que albergaban una eliminación de fondo de los exones 48-50 de longitud desconocida (Fig. 22A). La secuenciación de Sanger de esta banda de eliminación del ADN genómico de las células DMD tratadas reveló las uniones esperadas del intrón 44 y el intrón 55 inmediatamente adyacentes a los sitios diana del ARNgm (Fig. 22B). Después de la diferenciación de las células DMD tratadas, la eliminación esperada de los exones 45-55 se detectó en el transcripto de ARNm de distrofina y se verificó que era una fusión de los exones 44 y 56 mediante secuenciación de Sanger (Fig. 22C). La expresión de proteínas restaurada se observó mediante transferencia Western en las poblaciones de células clasificadas que contenían la eliminación inducida por CRISPR/Cas9 de los exones 45-55 del genoma y las transcripciones de ARNm resultantes (Fig. 22D). Estos datos demuestran que la edición múltiple CRISPR/Cas9 presenta un método universal único para restaurar el marco de lectura de distrofina en más del 60 % de las mutaciones de los pacientes con DMD.

25 Ejemplo 11

Trasplante de mioblastos corregidos en ratones inmunodeficientes

Un método prometedor para la terapia con DMD es corregir una población de células progenitoras musculares autólogas del paciente que pueden injertarse en el tejido del músculo esquelético del paciente para rescatar la expresión de distrofina. Para demostrar la capacidad de las células corregidas para expresar distrofina humana *in vivo*, una población de mioblastos DMD que fueron tratados con ARNgm CR1 y CR5, que flanquean el exón 51, se trasplantó y se clasificó para la expresión de GFP como antes (Fig. 19, Fig. 23). Después de 4 semanas, las fibras musculares positivas para espectrina humana, que se expresa tanto en células corregidas como no corregidas, se detectaron en criosecciones de tejido muscular inyectado (Fig. 24). Varias de estas fibras también fueron positivas para distrofina humana con expresión localizada en el sarcolema, demostrando la corrección genética funcional en estas células (Fig. 24, Fig. 25). No se observaron fibras positivas para distrofina humana en secciones de ratones inyectados con mioblastos DMD no tratados (Fig. 24, Fig. 25), lo que indica que las células modificadas con CRISPR/Cas9 fueron la fuente de expresión de distrofina humana.

40 Ejemplo 12

Análisis de citotoxicidad e inespecificidad

La citotoxicidad relativa del sistema CRISPR/Cas9 se evaluó en células humanas para ARNgm seleccionados adaptando un ensayo de retención de GFP basado en citometría de flujo como se describió anteriormente (Ousterout *et al.*, Mol Ther 21:1718-1726 (2013)). Se observó una citotoxicidad mínima para SpCas9 coexpresado con o sin ARNgm después de la transfección en células humanas (Fig. 26A). Hay herramientas disponibles públicamente para evaluar y priorizar la actividad potencial de CRISPR/Cas9 en loci inespecífico en función del sesgo posicional previsto de un desajuste determinado en la secuencia del protoespaciador de ARNgm y el número total de desajustes con el sitio diana previsto (Hsu *et al.*, Nat Biotechnol 31:827-826 (2013)). Este servidor web público se usó para predecir los sitios más probables inespecífico para los ARNgm usados para corregir el gen de la distrofina en este estudio (Tabla 4). Los diez principales sitios potenciales inespecífico se evaluaron mediante el ensayo de inspección en células HEK293T tratadas con SpCas9 y cassetes de expresión de ARNgm individuales para CR1, CR3, CR5, CR6 o CR36. CR1, CR3 y CR36 cada uno tenía uno de estos diez loci inespecífico previstos que demuestran niveles significativos de modificación génica (Tabla 4 y Fig. 27). De forma interesante, la secuencia CR3 inespecífica tenía una homología sustancial y frecuencias de modificación similares a la secuencia diana prevista (9,3 % en OT-1 frente a 13,3 % en el sitio previsto (Tabla 4 y Fig. 27). Notablemente, CR3-OT1 fue el único de estos tres sitios inespecíficos que mostró niveles significativos de actividad en las células hDMD clasificadas mediante el ensayo de inspección (Fig. 26B).

60 La actividad de la nucleasa en sitios inespecíficos puede provocar reordenamientos cromosómicos no deseados mediante re-ligación distal entre loci diana e inespecíficos escindidos en distintos cromosomas. Esto presenta una preocupación significativa para las estrategias de corrección de genes basadas en eliminación debido al mayor potencial de actividad inespecífica mediante el uso de dos o más nucleasas, tal como en la edición génica multiplex CRISPR/Cas9. Se probaron las posibles translocaciones mediante el uso de un ensayo de PCR genómica anidada altamente sensible para detectar translocaciones en los loci inespecíficos validados (Tabla 4) durante las estrategias

de edición CRISPR/Cas9 tanto simples como múltiples. Usando este ensayo, las translocaciones se detectaron fácilmente entre sitios específicos e inespecíficos en la línea celular modelo HEK293T que también muestra altos niveles de actividad inespecífica (Fig. 26C y Fig. 28A, 28B). La secuenciación de Sanger de los amplicones de PCR confirmó la identidad del evento de translocación previsto para cada par de cebadores (Figs. 29 y 30). Un subconjunto de las translocaciones detectadas en las células HEK293T también fueron detectables mediante PCR anidada en los mioblastos hDMD clasificados, aunque la señal fue considerablemente más débil y la identidad de la secuencia no se confirmó debido al bajo rendimiento de producto (Fig. 26D y Figs. 28A, 28C). No se detectaron translocaciones usando este ensayo en HEK293T o en células hDMD clasificadas tratadas con CR6 o CR6/CR36, respectivamente, (Fig. 28) que tenían niveles bajos de actividad inespecífica en CR6-OT3 solo en células HEK293T (Tabla 4). Estos resultados subrayan la importancia de seleccionar ARNgm altamente específicos, particularmente para aplicaciones de edición multiplex, y muestran que este enfoque puede beneficiarse de los esfuerzos continuos para mejorar la especificidad del sistema CRISPR/Cas9. Estos datos sugieren que los ARNgm seleccionados son capaces de corregir el gen de la distrofina sin una toxicidad significativa y con un solo sitio inespecífico fuertemente predicho con niveles detectables de actividad.

15

Ejemplo 13

Análisis

La edición genómica es una herramienta poderosa para corregir enfermedades genéticas y el reciente desarrollo del sistema CRISPR/Cas9 está acelerando drásticamente el progreso en esta área. La corrección de DMD, la enfermedad genética más común que actualmente tampoco tiene opciones terapéuticas aprobadas, quedó demostrada. Muchas terapias basadas en genes y células para la DMD se encuentran en desarrollo preclínico y en ensayos clínicos, y los métodos de edición genómica son compatibles con muchos de estos enfoques. Por ejemplo, la edición genómica puede combinarse con terapias basadas en células específicas del paciente para la DMD. El sistema CRISPR/Cas9 puede funcionar en células madre pluripotentes humanas y otras líneas celulares humanas, así como mioblastos esqueléticos humanos, como se muestra. De manera importante, la edición génica con CRISPR/Cas9 no abolió la capacidad miogénica de estas células, como lo demuestra la expresión eficiente de distrofina *in vitro* e *in vivo* después del trasplante en ratones inmunodeficientes. Por lo tanto, esta estrategia debería ser compatible con las terapias celulares para la DMD.

30

Adicionalmente, un conjunto enriquecido de células genéticamente corregidas demostró la expresión de distrofina humana *in vivo* tras el injerto en ratones inmunodeficientes. La edición génica CRISPR/Cas9 no tuvo efectos tóxicos significativos en los mioblastos humanos, como se observa mediante frecuencias de edición génica estables y la citotoxicidad mínima de varios ARNgm. Sin embargo, la actividad de edición génica se confirmó en tres de los 50 sitios inespecíficos previstos en cinco ARNgm y se detectaron translocaciones cromosómicas inducidas por CRISPR/Cas9 entre los sitios específicos e inespecíficos. La tecnología CRISPR/Cas9 es un método eficiente y versátil para corregir una fracción significativa de mutaciones de distrofina y puede servir como plataforma general para el tratamiento de enfermedades genéticas.

40

Adicionalmente, la transfección directa del ARNgm y el ARNm de Cas9, a diferencia del método de suministro basado en plásmidos usado aquí, puede usarse para aumentar la especificidad y la seguridad al reducir la duración de la expresión de Cas9 y eliminar la posibilidad de integración aleatoria del plásmido. Como alternativa, el suministro del sistema CRISPR/Cas9 directamente al músculo esquelético y/o cardíaco mediante vectores de suministro víricos, plasmídicos o de ARN puede usarse para la edición genómica *in vivo* y la traducción de este enfoque. El gran tamaño del gen Cas9 de *S. pyogenes* (~4,2 kilobases) presenta un desafío para su uso en vectores víricos adenoasociados de tamaño restringido. Sin embargo, los genes Cas9 de otras especies, tales como *N. meningitidis* y *S. thermophilus*, son lo suficientemente cortos como para empaquetar eficientemente casetes de expresión de Cas9 y ARNgm en vectores AAV individuales para aplicaciones de edición génica *in vivo*.

50

El sistema CRISPR/Cas9 permitió una modificación eficiente de casi el 90 % de las dianas probadas, consistente con otros informes de actividad robusta de este sistema en diversos loci. La robustez y la versatilidad de esta tecnología es un avance significativo hacia la creación a voluntad de edición génica específica del paciente. Bajos niveles de distrofina, incluyendo tan solo un 4 % de expresión de tipo silvestre, puede ser suficiente para mejorar la supervivencia, la función motora y la función cardíaca en un modelo de ratón. Los niveles de actividad CRISPR/Cas9 pueden ser suficientes para obtener un beneficio terapéutico.

60

El uso de multiplexación con CRISPR/Cas9 para eliminar exones también presenta un conjunto único de oportunidades y desafíos. Se realizó la eliminación de exones completos del genoma para restaurar la expresión de distrofina, en contraste con la restauración del marco de lectura del gen de la distrofina con pequeñas indeles generadas por la reparación del ADN basada en NHEJ tras la acción de una única nucleasa. El producto proteico del gen editado es predecible y ya está caracterizado en pacientes con distrofia muscular de Becker con la eliminación de origen natural, en contraste con las indeles aleatorias creadas por la acción intraexónica de una sola nucleasa que conducirá a la creación de nuevos epitopos a partir de cada evento de reparación del ADN. Adicionalmente, el producto resultante de las eliminaciones de exones conducirá a la restauración de la distrofina en cada evento de edición genética exitoso, mientras que modificar el gen con indeles aleatorias dentro de los exones solo restaurará el marco

de lectura en el tercio de los eventos de edición que conducen al marco de lectura correcto.

Ninguno de los ARN_{gm} analizados se asoció con efectos citotóxicos significativos en células humanas. Se identificaron tres sitios inespecíficos potenciales de un total de 50 sitios probados para los cinco ARN_{gm} usados para restaurar la expresión de distrofina. Adicionalmente, las translocaciones cromosómicas entre los sitios diana previstos y estos sitios inespecífico fueron detectables mediante ensayos de PCR anidados altamente sensibles en células HEK293T que expresan altos niveles de Cas9 y ARN_{gm}. Notablemente, la actividad inespecífico y las translocaciones identificadas en las células HEK293T, que es una línea celular inmortalizada y aneuploide que expresa niveles muy altos de Cas9 y ARN_{gm}, no se produjo en un nivel tan alto y en algunos casos fueron indetectables en los mioblastos hDMD. De manera importante, este nivel de especificidad puede ser tolerable dada la gravedad de la DMD, la falta de un aparente efecto citotóxico en las células humanas

Ejemplo 14

Una cápside AAV modificada por ingeniería genética, denominada SASTG (Fig. 34; SEQ ID NO: 436 y 437), se desarrolló para mejorar el tropismo del tejido del músculo cardíaco y esquelético (Piacentino *et al.* (2012) Human Gene Therapy 23:635-646). Una ZFN dirigida al locus Rosa26 ("Rosa26 ZFN"; Fig. 33; SEQ ID NO: 434 y 435) demostró ser altamente activa en células de ratón (Perez-Pinera *et al.* Nucleic Acids Research (2012) 40:3741-3752). Se diseñaron vectores AAV-SASTG que codifican la proteína Rosa26 ZFN y posteriormente se generaron y se purificaron mediante UNC Viral Vector Core. El ensayo de inspección (Guschin *et al.*, *Methods Mol Biol* 649, 247-256 (2010)) se usó para demostrar la mutagénesis NHEJ en el locus Rosa26 después del suministro de ZFN Rosa26 de AAV-SASTG en mioblastos C2C12 cultivados que estaban ciclando activamente o forzados a diferenciarse mediante la eliminación de suero (no mostrado).

Para verificar que el músculo esquelético posmitótico adulto fue atacado eficientemente por Rosa26 ZFN después de la transferencia de AAV, los vectores AAV-SASTG que codifican Rosa26 ZFN se inyectaron directamente en el músculo tibial anterior (TA) de ratones C57BL6/J de 6 semanas de edad con títulos de genomas de vector le10 (vg) o 2,5e10 vg por músculo. Los ratones se sacrificaron 4 semanas después de la inyección y los músculos TA se recogieron y dividieron en varios fragmentos para la extracción y análisis del ADN genómico (Fig. 31). El ADN genómico se amplificó por PCR y se sometió al ensayo de inspección para detectar mutaciones NHEJ características de la mutagénesis ZFN en el sitio diana Rosa26 (Fig. 32). La Fig. 32 muestra el análisis de inspección de las actividades de Rosa26 ZFN en músculo esquelético *in vitro* e *in vivo* después del suministro de AAV-SASTG-ROSA. C2C12 en proliferación se transdujeron con la cantidad indicada de virus y se recogieron 4 días después de la infección (Fig. 32(a)). C2C12 se incubaron en medio de diferenciación durante 5 días y después se transdujeron con la cantidad indicada de virus AAV-SASTG-ROSA en placas de 24 pocillos (Fig. 32(b)). Las muestras se recogieron 10 días después de la transducción. La cantidad indicada de AAV-SASTG-ROSA se inyectó directamente en el tibial anterior de ratones C57BL6/J y los músculos se recogieron 4 semanas después de la infección. Los músculos TA recogidos se dividieron en 8 trozos separados para el análisis de ADN genómico, cada uno se muestra en un carril separado (Fig. 32(c)). Notablemente, se detectaron altos niveles de modificación génica en todos los fragmentos a la dosis más alta (2,5e10 vg).

Ejemplo 15

Construcciones AAV-CRISPR para dirigirse a genes mutantes de distrofina

Las construcciones de AAV están diseñadas para corregir terapéuticamente las mutaciones del gen de la distrofina que provocan la distrofia muscular de Duchenne y la degeneración del músculo esquelético y cardíaco. Los sistemas CRISPR/Cas9 pueden suministrarse usando el AAV para restaurar el marco de lectura de distrofina eliminando el exón 51, eliminando los exones 45-55, alterando los sitios donantes o aceptores de corte y empalme o creando desplazamientos de marco dentro del exón 51 (Ousterout *et al.*, Molecular Therapy 2013) para restaurar el marco de lectura de distrofina y la expresión de proteínas. El sistema CRISPR/Cas9 incluirá un Cas9 que tiene una secuencia de SEQ ID NO: 64 o 114 (véanse las Fig. 40 y 41). Los ARN_g que podrían combinarse con estos Cas9, dirigiéndose a sus respectivas secuencias PAM, se proporcionan (véanse las Fig. 40 y 41; véanse también las Tablas 2 y 3).

Ejemplo 16

Generación de Neuronas inducidas (iN)

La generación de neuronas inducidas (iN) a partir de otros linajes celulares tiene aplicaciones potenciales en la medicina regenerativa y el estudio de enfermedades neurológicas. La conversión directa de fibroblastos embrionarios de ratón (MEF) en células neuronales funcionales puede producirse mediante el suministro de un cóctel de tres factores de transcripción neuronal *BRN2*, *ASCL1* y *MYT1L* (Factores BAM, Fig. 48). Otros métodos pueden incluir factores adicionales para inducir diversos subtipos neuronales. Estos experimentos requieren el suministro ectópico de factores de transcripción y la activación de los loci endógenos correspondientes para mantener el fenotipo neuronal. El sistema CRISPR/Cas9 fue diseñado como un factor de transcripción versátil para activar genes endógenos en células de mamíferos con la capacidad de dirigirse a cualquier promotor en el genoma a través de un mecanismo

guiado por ARN (Fig. 49A, 49B).

Materiales y métodos. Se usó el factor de transcripción CRISPR/Cas9 para activar los genes endógenos que codifican ASCL1 y BRN2 para reprogramar directamente MEF en neuronas inducidas funcionales.

5 **Cultivo celular:** Los MEF se sembraron en placas TCPS de 24 pocillos o en cubreobjetos recubiertos con poli-D-lisina/laminina. Después de la transducción de dCas9-VP64 y la transfección de los ARNg (véanse las Tablas 10 y 11 para ver las secuencias de ARNg), las células se cultivaron en medio MEF (Adler *et al.* Mol Ther Nucleic Acids 1:e32 (2012)) durante 24 h y después se transfirieron a un medio de inducción neural N3 (Vierbuchen *et al.* Nature 463:1035-1041 (2010)) durante la duración del experimento (Fig. 49B).

10 **Tabla 10 ARNg para ASCL1 de ratón (CR13):**

Oligo (5' a 3')	saliente 5'		Secuencia diana ASCL1	SEQ ID NO
CR13-1_S:	cacc	G	CAGCCGCTCGCTGCAGCAG (SEQ ID NO: 468)	492
CR13-1_AS:	AAAC		CTGCTGCAGCGAGCGGGCTG (SEQ ID NO: 469)	C 493
CR13-2_S:	cacc	G	GCTGGGTGTCCCATTGAAA (SEQ ID NO: 470)	494
CR13-2_AS:	AAAC		TTTCAATGGGACACCCAGC (SEQ ID NO: 471)	C 495
CR13-3_S:	cacc	G	GT TT ATT CAGCCGGGAGTC (SEQ ID NO: 472)	496
CR13-3_AS:	AAAC		GACTCCCCGCTGAATAAAC (SEQ ID NO: 473)	C 497
CR13-4_S:	cacc	G	TGGAGAGTTGCAAGGAGC (SEQ ID NO: 474)	498
CR13-4_AS:	AAAC		GCTCCTTGCAA ACTCTCCA (SEQ ID NO: 475)	C 499
CR13-5_S:	cacc	G	CCCTCCAGACTTTCCACCT (SEQ ID NO: 476)	500
CR13-5_AS:	AAAC		AGGTGGAAAGTCTGGAGGG (SEQ ID NO: 477)	C 501
CR13-6_S:	cacc	G	AATTTTCTTCCAAGTTCTC (SEQ ID NO: 478)	502
CR13-6_AS:	AAAC		GAGAACTT GGAAGAAAATT (SEQ ID NO: 479)	C 503
CR13-7_S:	cacc	G	CTGCGGAGAGAAGAAAGGG (SEQ ID NO: 480)	504
CR13-7_AS:	AAAC		CCCTTCTCTCTCCGCAG (SEQ ID NO: 481)	C 505
CR13-8_S:	cacc	G	AGAGCCACCCCTGGCTCC (SEQ ID NO: 482)	506
CR13-8_AS:	AAAC		GGAGCCAGGGGGTGGCTCT (SEQ ID NO: 483)	C 507
CR13-9_S:	cacc	G	cgaagccaaccgcggcgagg (SEQ ID NO: 484)	508
CR13-9_AS:	AAAC		cccgccgcegggttggcttcg (SEQ ID NO: 485)	C 509
CR13-10_S:	cacc	G	agagggaagacgategeccc (SEQ ID NO: 486)	510
CR13-10_AS:	AAAC		ggg c gatcgttccctct (SEQ ID NO: 487)	C 511
CR13-11_S:	cacc	G	cccctttaacttccctcg (SEQ ID NO: 488)	512
CR13-11_AS:	AAAC		cggaggaaagttaagggg (SEQ ID NO: 489)	C 513
CR13-12_S:	cacc	G	gcagccccgc ttc tcaa (SEQ ID NO: 490)	514
CR13-12_AS:	AAAC		ttgaaggaagcggggctgc (SEQ ID NO: 491)	C 515

Tabla 11 ARNg para BRN2 de ratón (CR16):

Oligo (5' a 3')	saliente 5'		Secuencia diana BRN2	SEQ ID NO
CR16-1_S:	cacc	G	CGAGAGCGAGAGGGAGGGAG (SEQ ID NO: 516)	540
CR16-1_AS:	AAAC		CTCCCTCCTCTCGCTCTCG (SEQ ID NO: 517)	C 541
CR16-2_S:	cacc	G	GAGAGAGCTTGAGAGCGCG (SEQ ID NO: 518)	542
CR16-2_AS:	AAAC		CGCGCTCTCAAGCTCTCTC (SEQ ID NO: 519)	C 543
CR16-3_S:	cacc	G	GGTGGAGGGGGCGGGGCC (SEQ ID NO: 520)	544
CR16-3_AS:	AAAC		GGGCCCCGCCCCCTCCACC (SEQ ID NO: 521)	C 545
CR16-4_S:	cacc	G	GGTATCCACGTAAATCAA (SEQ ID NO: 522)	546
CR16-4_AS:	AAAC		TTTGATTTACGTGGATACC (SEQ ID NO: 523)	C 547

CR16-5_S:	cacc	G	CCAATCACTGGCTCCGGTC (SEQ ID NO: 524)		548
CR16-5_AS:	AAAC		GACCGGAGGCCAGTGATTGG (SEQ ID NO: 525)	C	549
CR16-6_S:	cacc	G	GGCGCCCGAGGGAAAGAAGA (SEQ ID NO: 526)		550
CR16-6_AS:	AAAC		TCTTCTCCCTCGGGCGCC (SEQ ID NO: 527)	C	551
CR16-7_S:	cacc	G	GGGTGGGGGTACCAGAGGA (SEQ ID NO: 528)		552
CR16-7_AS:	AAAC		TCCTCTGGTACCCCCACCC (SEQ ID NO: 529)	C	553
CR16-8_S:	cacc	G	CCGGGGACAGAAGAGAGGG (SEQ ID NO: 530)		554
CR16-8_AS:	AAAC		CCCTCTCTCTGTCCCCGG (SEQ ID NO: 531)	C	555
CR16-9_S:	cacc	G	gagagagatgggagaagc (SEQ ID NO: 532)		556
CR16-9_AS:	AAAC		gtttctccactctcttc (SEQ ID NO: 533)	C	557
CR16-10_S:	cacc	G	aaagtaactgtcaaatgcg (SEQ ID NO: 534)		558
CR16-10_AS:	AAAC		cgcatttgacagttacttt (SEQ ID NO: 535)	C	559
CR16-11_S:	cacc	G	ttaaccagagcgcccagtc (SEQ ID NO: 536)		560
CR16-11_AS:	AAAC		gactgggcgcctggtaa (SEQ ID NO: 537)	C	561
CR16-12_S:	cacc	G	cgtcgagactgcggctag (SEQ ID NO: 538)		562
CR16-12_AS:	AAAC		ctagcggcagctccgacg (SEQ ID NO: 539)	C	563

qRT-PCR e IF: La activación de *ASCL1* endógeno se evaluó mediante qRT-PCR e inmunofluorescencia en MEF el día 3 después del suministro de dCas9-VP64 y ARNg, ADNc de *ASCL1* o un vector de control negativo que codifica luciferasa. La generación de iN se evaluó mediante co-tinción de *TUJ1* y *MAP2* e identificación de células con morfología neuronal y procesos extendidos.

Indicadores de células vivas: Después de 7-8 días en medio N3, las MEF cultivadas en cubreobjetos recubiertos con poli-D-lisina/laminina se transdujeron con virus que portaban indicadores hSyn-RFP y MAP2-GCaMP5 para identificar las iN más maduras para la caracterización funcional mediante imágenes de calcio y electrofisiología (Fig. 49B).

Resultados. dCas9-VP64 y ARNg dirigidos al promotor de *ASCL1* activaron el gen endógeno en las MEF. El cosuministro de 8 ARNg activó el gen endógeno 400 veces, un aumento significativo ($p<0,05$) con respecto a la activación de 100 veces inducida por el cosuministro de 4 ARNg (Fig. 50A). La proteína *Ascl1* localizada en el núcleo se detectó por inmunofluorescencia en MEF. La expresión de *Ascl1* ectópica produjo más proteína *Ascl1* que dCas9-VP64 con cualquiera de los cócteles de ARNg, pero no activó el locus endógeno el día 3 (Fig. 50A, 50B). Las células copositivas para *TUJ1* y *MAP2* con procesos extendidos se identificaron mediante inmunofluorescencia después de 13 días en medio neurogénico después del suministro de dCas9-VP64 y ARNg dirigidos a los promotores *ASCL1* y *BRN2* (Fig. 4A primera fila). Se identificó un número similar de células copositivas para *TUJ1* y *MAP2* con expresión ectópica de los factores BAM (Fig. 51A, segunda fila). Las células con una morfología neuronal que expresaban el indicador hSyn-RFP fueron visibles en cultivo ya el día 11 en medio neurogénico (Fig. 51B). Una célula que expresa el indicador de calcio MAP2-GCaMP5 exhibió despolarización inducida por KCl detectada con un microscopio fluorescente (Fig. 52A, 52B).

La conversión directa de fibroblastos embrionarios de ratón en células copositivas para *TUJ1* y *MAP2* con una morfología neuronal se logró mediante la activación de células endógenas *BRN2* y *ASCL1* por factores de transcripción basados en CRISPR/Cas9. Aunque dCas9-VP64 produce menos proteínas que la expresión ectópica de *ASCL1* (Fig. 50B), la generación de células de tipo neuronal es similar. La activación de los loci endógenos puede inducir una cascada de eventos de reprogramación que no es mecánicamente idéntica a la generada con la expresión ectópica.

dCas9-VP64 pudo penetrar la heterocromatina y activó genes endógenos silenciados de manera estable, una característica de solo un subconjunto de factores de transcripción "pioneros". Como resultado, convirtiendo linaje celular con factores de transcripción CRISPR/Cas9 pueden superar mejor las barreras epigenéticas para la reprogramación que la expresión ectópica de factores de transcripción, particularmente en tipos de células difíciles de reprogramar, tales como células humanas adultas. Esto puede tener importancia clínica en el campo de la medicina regenerativa, ya que a menudo se desea usar fuentes autólogas en terapias de reemplazo celular.

Ejemplo 17

Modificación por ingeniería genómica basada en Multiplex CRISPR/Cas9: materiales y métodos

Construcciones de plásmidos. Se usaron los casetes de expresión para el ARNm de *S. pyogenes* y la nucleasa Cas9 (hCas9) optimizada con codones humanos, como se describió anteriormente. Los promotores adicionales para

mU6 (Ohshima *et al.*, Nucleic Acids Res 9:5145-5158 (1981)), HI (Myslinski *et al.*, Nucleic Acids Res 29:2502-2509 (2001)) y 7SK (Murphy *et al.*, Cell 51:81-87 (1987)) promotores pol-III se sintetizaron usando GeneBlocks (IDT) y se clonaron en lugar del casete de expresión de ARNgm hU6. Se clonó un GeneBlock (IDT) en el extremo 3' de la secuencia codificante de Cas9 para fusionar un péptido de omisión T2A y el gen ϵ GFP inmediatamente después de

5 Cas9 para monitorizar la expresión del vector. La región codificante de hCas9-T2A-GFP (SEQ ID NO: 145) se transfirió después a un vector de expresión lentivírico que contenía el promotor de ubiquitina C humana (hUbC) para impulsar la expresión de hCas9-T2A-GFP, así como sitios de restricción para facilitar la clonación Golden Gate de cassetes de expresión de ARNgm inmediatamente en dirección 5' del promotor hUbC (Fig. 42A).

10 **Protocolo para el ensamblaje de vectores lentivíricos personalizados.** El ensamblaje de vectores lentivíricos personalizados que expresan hasta cuatro ARNgm de elección y Cas9 activa, dCas9 o dCas9-VP64 se logró en menos de cinco días. El método de clonación usó la clonación Golden Gate y las enzimas de restricción tipo IIS que se escinden fuera de su secuencia de reconocimiento para crear salientes únicos. El ensamblaje Golden Gate aceleró la clonación ya que los cuatro cassetes de expresión se ligaron al vector lentivírico final en una sola etapa. El vector 15 lentivírico expresó Cas9 activa, cCas9 o dCas9-VP64 además de uno, dos, tres o cuatro ARNgm expresados a partir de promotores independientes.

20 Etapa 1: Los oligos monocatenarios que contenían cada protoespaciador de 20 pb se hibridaron de tal manera para crear extremos pegajosos y se ligaron en el vector promotor pZDonor deseado. Se pidieron dos oligos monocatenarios para cada diana genómica deseada. Para hibridar los oligos complementarios, mezclar 8 μ l de oligo sentido + 8 μ l de oligo antisentido (ambos a 10 mM) + 2 μ l de tampón ligasa 10X. Los oligos se hibridan y se vuelven a hibridar en un termociclador con el programa: 96 °C durante 300 segundos, seguido de 85 °C durante 20 segundos, 75 °C durante 20 segundos, 65 °C durante 20 segundos, 55 °C durante 20 segundos, 45 °C durante 20 segundos, 35 °C durante 20 segundos y 25 °C durante 20 segundos con una velocidad de -0,3 °C/segundo entre etapas. Para fosforilar los 25 extremos pegajosos, añadir 1 μ l de ATP 25 mM + 1 microlitro de Polinucleótido Quinasa de T4 (NEB) e incubar a 37 °C durante 60 minutos seguido de 65 °C durante 20 minutos para inactivar por calor la enzima. Cada protoespaciador se ligó al vector de expresión deseado usando ADN ligasa T4 (NEB) incubada a 16 °C durante 60 minutos usando 50 ng de vector y 1 μ l de oligonucleótidos hibridados en un volumen de reacción de 10 μ l de acuerdo con las instrucciones del fabricante. Se transformaron cinco microlitros de cada ligación en bacterias químicamente competentes XL1 blue 30 (Agilent) siguiendo las instrucciones del fabricante. Transformación de placas en placas de agar LB que contienen 50 μ g/ml de kanamicina (Sigma) e incubación durante la noche a 37 °C. En la presente experiencia, >90 % de las colonias contendrán el producto de ligación deseado. Se realizó la secuenciación usando el cebador de secuenciación estándar inverso MI3 para validar cada construcción de ARNgm final antes de pasar a la etapa 2.

35 *Etapa 2: Construcción de los cuatro cassetes de ARNg promotor en un vector de destino lentivírico usando el ensamblaje Golden Gate.* Después de completar la etapa 1, hay cuatro plásmidos independientes cada uno de los cuales expresa un ARNgm diferente de un promotor diferente. Para ensamblar las cuatro construcciones diferentes de promotor-ARNgm en el vector de destino deseado, combinar 200 ng de cada plásmido de expresión de ARNgm y vector de destino lentivírico deseado con 1 μ l de ADN ligasa T4 (NEB), 1 μ l de BsmBI FastDigest (Fisher Scientific) y 40 2 μ l de tampón de ligasa T4 10X (NEB) en un volumen de reacción de 20 μ l, incubar la reacción como sigue: 37 °C durante 10 minutos, 16 °C durante 15 minutos, 37 °C durante 30 minutos, 80 °C durante 5 minutos. Transformar 5 μ l de reacción de ligación en células SURE 2 químicamente competentes (Agilent) siguiendo las instrucciones del fabricante. Las transformaciones en placas se realizan en placas de agar LB que contienen 100 μ g/ml de ampicilina y 45 se incuban durante la noche a 37 °C. Opcionalmente, las colonias pueden examinarse mediante detección azul/blanca basada en lacZ usando 1PTG y X-gal; sin embargo, en la experiencia de los presentes inventores, >90 % de los transformantes contienen el producto de ligación adecuado. Debido a las repeticiones invertidas formadas por los cassetes de expresión de ARNgm opuestos, las construcciones finales pueden ser inestables y, por lo tanto, se recomienda mantener estos plásmidos en la línea celular SURE 2 y examinar el plásmido final con una prueba de PCR usando el cebador sentido 5'-TCGGGTTTATTACAGGGACAGCAG-3' (SEQ ID NO:464) y el cebador antisentido 5'-50 TCTAAGGCCGAGTCTTATGAGCAG-3' (SEQ ID NO:465). Estos cebadores se amplifican en las cuatro regiones promotoras de ARNg. Debido a la naturaleza repetitiva, debe observarse un patrón de bandas distinto con el producto más grande de aproximadamente 1800 pb de tamaño.

55 **Cultivo celular y transfección.** Las células HEK293T se obtuvieron del American Tissue Collection Center (ATCC, Manassas, VA) a través de las instalaciones del Centro Oncológico de la Universidad de Duke y se mantuvieron en DMEM suplementado con FBS al 10 % y penicilina y estreptomicina al 1 %. Los fibroblastos dérmicos humanos primarios (ID de catálogo: GM03348) se obtuvieron del Instituto Coriell (Camden, Nueva Jersey) y se mantuvieron en DMEM suplementado con FBS al 10 % y penicilina y estreptomicina al 1 %. Todas las células se cultivaron a 37 °C con CO₂ al 5 %. Las células HEK293T se transfecaron con Lipofectamine 2000 (Life Technologies) con 200 ng de cada vector de expresión de ARNgm (800 ng de pDNA total) de acuerdo con el protocolo del fabricante en placas de 24 pocillos.

60 **Producción y transducción vírica.** Todos los vectores lentivíricos usados en este estudio son de segunda generación y se produjeron usando métodos de producción vírica convencionales. Brevemente, se sembraron 3,5 millones de HEK293T por placa de 10 cm. El día siguiente, las células se transfecaron mediante el método de transfección con fosfato cálcico con 20 μ g de vector de transferencia, 6 μ g de pMD2G y 10 μ g de psPAX2. El medio se cambió 12-

14 horas después de la transfección. El sobrenadante vírico se recogió 24 y 48 horas después de este cambio de medio, se pasó a través de un filtro de 0,45 micrómetros y se agrupó. Para la transducción, el medio celular se reemplazó con sobrenadante vírico suplementado con 4 µg/ml de polibreno. El sobrenadante vírico se intercambió por medio nuevo 12-24 horas después.

5 **PCR de transcripción inversa.** El ARN se aisló usando el kit de aislamiento de ARN miRNeasy Mini (Qiagen). La digestión con DNase se realizó usando el kit ADN-free (Applied Biosystems). La síntesis de ADNc se realizó usando el kit de síntesis de ADNc Superscript VILO (Invitrogen). El ADNc se amplificó por PCR usando ADN polimerasa Taq (NEB) y el producto resultante se ejecutó en geles de agarosa TAE. Las imágenes se capturaron usando un sistema ChemiDoc XRS y se procesaron usando el software ImageLab (Bio-Rad).

10 **PCR cuantitativa en tiempo real.** El ARN se aisló usando el kit de aislamiento de ARN RNeasy Plus (Qiagen). La síntesis de ADNc se realizó usando el kit de síntesis de ADNc Superscript VILO (Invitrogen). La PCR en tiempo real usando PerfeCTa SYBR Green FastMix (Quanta Biosciences) se realizó con el sistema de detección de PCR en tiempo real CFX96 (Bio-Rad). La especificidad del cebador se confirmó mediante electroforesis en gel de agarosa y análisis de la curva de fusión. Se calcularon las eficiencias de reacción en el intervalo dinámico apropiado para garantizar la linealidad de la curva patrón. Los resultados se expresan como expresión de ARNm multiplicada por el gen de interés normalizado a la expresión de β -actina usando el método $\Delta\Delta Ct$. Los valores informados son la media y S.E.M. de dos experimentos independientes ($n = 2$) donde se promediaron las réplicas técnicas para cada experimento.

15 **Transferencia Western.** Las células se lisaron en tampón RIPA (Sigma) suplementado con un cóctel inhibidor de proteasa (Sigma). La concentración de proteínas se midió usando el reactivo de ensayo de proteínas BCA (Thermo Scientific) y el lector de microplacas multimodo BioTek Synergy 2. Los lisados se mezclaron con tampón de carga y se hirvieron durante 5 minutos; se procesaron 25 µg de proteína en geles de poliacrilamida NuPage al 10 % Bis-Tris Gel (Bio-Rad) y se transfirieron a membranas de nitrocelulosa. La unión de anticuerpos no específicos se bloqueó con TBST (Tris 50 mM, NaCl 150 mM y Tween-20 al 0,1 %) con leche desnatación al 5 % durante 1 hora a temperatura ambiente. Las membranas se incubaron con anticuerpos primarios: (anti-MyoD (1:250, Santa Cruz Sc-32758) en BSA al 5 % en TBST durante la noche a 4 °C; anti-Miogenina (1:250, Santa Cruz Sc-12732) en BSA al 5 % durante la noche a 4 °C; anti-FLAG-HRP (1:1000, Cell Signaling 2044) en leche al 5 % en TBST durante 60 min a temperatura ambiente; anti-GAPDH (1:5000, Cell Signaling, clon 1400) en leche al 5 % en TBST durante 30 min a temperatura ambiente. Las membranas se lavaron después tres veces con TBST durante 15 minutos en total. Las membranas se incubaron con anticuerpo anti-conejo conjugado con HRP (Sigma, A 6154) o anticuerpo anti-ratón conjugado con HRP (Santa Cruz, SC-2005) diluido 1:5000 durante 30 min y se lavaron con TBST tres veces durante 15 min cada una. Las membranas se visualizaron usando el kit de quimioluminiscencia ImmunStar WesternC (Bio-Rad) y las imágenes se capturaron usando un sistema ChemiDoc XRS y se procesaron usando el software ImageLab (Bio-Rad).

20 **Cuantificación Cel-I de la modificación de genes endógenos.** Las lesiones de nucleasa CRISPR/Cas9 en el sitio diana endógeno se cuantificaron usando el ensayo de nucleasa de inspección, que puede detectar mutaciones características de NHEJ mediada por nucleasas. Después de la transfección o transducción, las células se incubaron durante 3-10 días a 37 °C y se extrajo el ADN genómico usando el kit DNeasy Blood and Tissue (Qiagen). El locus diana se amplificó mediante 35 ciclos de PCR con el kit AccuPrime High Fidelity PCR (Invitrogen). Los productos de PCR resultantes se fundieron aleatoriamente y se volvieron a hibridar en un termociclador con el programa: 95 °C durante 240 segundos, seguido de 85 °C durante 60 segundos, 75 °C durante 60 segundos, 65 °C durante 60 segundos, 55 °C durante 60 segundos, 45 °C durante 60 segundos, 35 °C durante 60 segundos y 25 °C durante 60 segundos con una velocidad de ~0,3 °C/segundo entre etapas. Después de la rehibridación, se mezclaron 8 µl de producto de PCR con 1 µl de Nucleasa S de inspección y 1 µl de Potenciador S (Transgenomic, Omaha, NE) y se incubaron a 42 °C durante 1 h. Después de la incubación, se cargaron 6 µl del producto de digestión en un gel de poliacrilamida TBE al 10 % y se ejecutó a 200 V durante 30 minutos. Los geles se tiñeron con bromuro de etidio y se cuantificaron usando ImageLab (Bio-Rad) mediante densitometría (Perez-Pinera *et al.*, Nucleic Acids Res 40:3741-3751 (2012)).

25 **Análisis estadístico.** Se compilaron al menos dos experimentos independientes como medias y errores estándar de la media. Los efectos se evaluaron con ANOVA multivariado y la prueba post hoc de Dunnett usando JMP 10 Pro.

55 **Ejemplo 18**

Desarrollo de un único vector lentivírico para aplicaciones multiplex CRISPR/Cas9

60 Una limitación de los sistemas actuales de edición génica CRISPR/Cas9, particularmente sistemas transactivadores, es el suministro simultáneo y eficiente de múltiples ARNm y proteína Cas9 para aplicaciones de activación de genes sinérgica y edición génica múltiple, especialmente en tipos de células difíciles de transfectar. Para superar esta limitación, los presentes inventores desarrollaron un único vector lentivírico que expresa eficientemente Cas9 y hasta cuatro ARNm. Para maximizar la eficiencia de expresión de cada ARNm, este vector expresa cuatro ARNm de cuatro promotores pol III independientes (promotor IJ6 humano, promotor del ratón U6, 7SK y H1). Los presentes inventores validaron la expresión de ARNm de cada promotor usando RT-PCR de punto final para detectar un ARNm

dirigido al locus AAVS1 (Fig. 42 A). Para probar la actividad de cada construcción de expresión de ARNgm, los presentes inventores cotransfectaron cada construcción promotora que expresa un ARNgm dirigido a AAVS1 de forma independiente con una construcción de expresión activa de Cas9 en células HEK293T humanas. Notablemente, los presentes inventores detectaron niveles consistentes y altos de modificación génica en el locus diana para cada

5 ARNgm que son comparables a una nucleasa con dedos de cinc bien caracterizada con alta actividad en el locus AAVS1 (Fig. 42B). Adicionalmente, el suministro lentivírico de diferentes construcciones basadas en Cas9, incluyendo una nucleasa Cas9 activa, Cas9 muerta y Cas9 muerta fusionada al dominio transactivador VP64, dio como resultado la expresión de la proteína Cas9 de longitud completa en células HEK293T según lo determinado mediante transferencia Western (Fig. 42C).

10 Usando estos componentes, los presentes inventores desarrollaron un método de clonación Golden Gate para facilitar la clonación rápida y eficiente de múltiples casetes de expresión de ARNgm en un único vector lentivírico que expresa el efector Cas9 deseado (Fig. 43). En la primera etapa, los oligonucleótidos que codifican secuencias protoespaciadoras de ARNgm se clonian de forma independiente en diferentes vectores de expresión, cada uno con 15 un promotor distinto que impulsa la expresión de ARNgm. En la segundo etapa, cada construcción de expresión de ARNgm se subclona en un vector de expresión lentivírica Cas9 de elección mediante el ensamblaje Golden Gate. Esta estrategia permite una clonación sólida y rápida de hasta cuatro ARNgm en un único vector lentivírico para aplicaciones de activación o edición génica. Para expresar menos de cuatro ARNgm, se clona una secuencia terminadora PoliT en dirección 3' de los promotores no usados para evitar la transcripción de los promotores no usados. Cada vector 20 coexpresa la elección Cas9 con eGFP a través de un péptido de omisión 2A para permitir la clasificación del flujo activado por fluorescencia y el enriquecimiento de células con una alta multiplicidad de infección. Finalmente, toda la región que contiene los casetes de expresión de ARNgm y Cas9 está flanqueada por sitios loxP para mediar la eliminación mediante escisión Cre-lox.

25 Ejemplo 19

Validación de un único vector de expresión lentivírica ARNgm/Cas9 para ingeniería genómica múltiple

30 Para validar la actividad independiente de cada ARNgm, los presentes inventores ensamblaron un único vector lentivírico que expresa Cas9 activo y cuatro ARNgm, cada uno de ellos dirigido a un loci independiente (Fig. 44A). Como vectores de control, los presentes inventores ensamblaron construcciones que expresan solo un ARNgm junto con protoespaciadores poliT en las otras tres posiciones. Los presentes inventores transdijeron HEK293T y fibroblastos primarios con vectores lentivíricos que expresan los ARNgm indicados y monitorizaron las frecuencias de 35 modificación génica a los 7 o 10 días después de la transducción, respectivamente (Fig. 44B). En ambos tipos de células, el único vector lentivírico medió una edición múltiple altamente eficiente en los cuatro loci (Fig. 44B). De forma interesante, la expresión de los cuatro ARNgm juntos dio como resultado frecuencias de modificación más altas que un solo ARNgm solo en 3 de 4 loci en fibroblastos (Fig. 44B). Los presentes inventores observaron una edición 40 génica múltiple eficiente en fibroblastos, que son convencionalmente un tipo de célula difícil de transfectar. Estos datos demuestran que un solo lentivirus puede expresar cuatro ARNgm activos de manera eficiente y que esta plataforma lentivírica puede usarse para dirigirse a cuatro loci distintos para la edición génica múltiple CRISPR/Cas9.

Ejemplo 20

Activación transitoria de genes guiada por ARN en líneas celulares que expresan de manera estable un transactivador lentivírico basado en Cas9

45 A continuación, los presentes inventores estaban interesados en desarrollar un sistema que permita la activación transitoria de genes mediante la transfección de ARNgm en líneas celulares modelo que expresen Cas9 de manera estable. Las HEK293T se transdijeron con diferentes Cas9-T2A-GFP y la expresión de GFP se controló mediante 50 citometría de flujo. Después de un pasaje normal cada 2-3 días, cada línea celular exhibió una expresión estable de GFP hasta 35 días después de la transducción. Las HEK293T transducidas se transfecaron después con de una a cuatro construcciones de expresión de ARNgm separadas dirigidas al promotor IL1RN o HBG1. La transfección transitoria de estas construcciones de ARNgm en líneas celulares estables que expresan dCas9-VP64 dio como 55 resultado una activación de genes endógenos ajustables (Fig. 45A, 45B). La activación génica después de la transfección transitoria de construcciones de ARNgm en células que expresan dCas9-VP64 alcanzó un nivel máximo de activación aproximadamente 3-6 después de la transfección y cayó a niveles indetectables 20 días después de la transfección (Figs. 45C, 45D). Adicionalmente, los presentes inventores fueron capaces de reactivar cada gen mediante una segunda transfección de las cuatro construcciones de ARNgm dirigidas a cada promotor, aunque los niveles de activación fueron significativamente más bajos que los observados en la primera transfección (Fig. 45C, 60 45D). Esta reducción de la actividad después de la segunda transfección puede deberse a una expresión reducida del vector o al crecimiento competitivo de las células no transducidas. A pesar de esto, estos datos demuestran que la Cas9 lentivírica combinado con el suministro de ARNgm transitorio puede usarse como un sistema versátil para activar y reactivar de forma ajustable y transitoria genes diana en una línea celular transducida de forma estable con Cas9.

65 Ejemplo 21

Activación génica estable en células HEK293T usando un único vector de expresión transactivador ARNgm/Cas9 lentivírico

El suministro lentivírico puede permitir una activación génica estable a largo plazo mediante transactivación CRISPR/Cas9. Para probar esto, se transdijeron HEK293T usando un único vector lentivírico que codifica dCas9-VP64 y de uno a cuatro cassetes de expresión de ARNgm. De manera similar a los resultados de transfección transitoria (Fig. 45), se pudo activar de manera robusta y ajustable la expresión de los genes endógenos IL1RN y HBG1 (Figs. 46A, 46B). La activación génica inducida por la cotransfección de células HEK293T con dCas9-VP64 y cuatro ARNgm dirigidos a los promotores IL1RN y HBG1 alcanzó su punto máximo entre tres y cinco días después de la transfección y la expresión génica volvió a los niveles de fondo entre 15-20 días después de la transfección (Fig. 4c). Por el contrario, el suministro lentivírico de dCas9-VP64 y los mismos cuatro ARNgm dirigidos a IL1RN o HBG1 indujeron una activación génica sostenida durante más de 20 días después de la transducción (Fig. 46C, 46D). Por lo tanto, el suministro lentivírico individual de transactivadores dCas9-VP64 múltiple es una plataforma útil para regular positivamente de manera eficiente y estable genes endógenos diana.

15

Ejemplo 22

dCas9-KRAB: Direccionamiento al potenciador HS2

El potenciador HS2 es un elemento regulador distal bien caracterizado y necesario para la activación del locus del gen de la globina. Se suministraron dCas9-KRAB con ARNg dirigidos al potenciador HS2 para determinar si este sistema reprimiría la expresión de γ , ϵ - y β -globina en la línea celular de leucemia eritroide humana K562 (Fig. 54). Se creó un panel de ARNg dirigidos a diferentes sitios a lo largo de la región central del potenciador HS2 (SEQ ID NO: 467). Véase la Tabla 12.

25

Tabla 12 Secuencias diana de ARNg de HS2

N.º de Cr	Protoespaciador	Secuencia completa de ARNg
1	gagacacacagaaaatgtAAC (SEQ ID NO: 564)	gagacacacagaaaatgtAACttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 585)
2	ggtggggcactgaccgcac (SEQ ID NO: 565)	ggtggggcactgaccgcacttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 586)
3	ctagagtgtactccatc (SEQ ID NO: 566)	ctagagtgtactccatcgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 587)
4	gactaaaactccaccta (SEQ ID NO: 567)	gactaaaactccacctaaggttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 588)
5	aatatgtcacatttgttc (SEQ ID NO: 568)	aatatgtcacatttgttcgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 589)
6	ggactatggaggactaa (SEQ ID NO: 569)	ggactatggaggactaaaggttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 590)
7	gctcatgtggactatgg (SEQ ID NO: 570)	gctcatgtggactatgggtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 591)
8	gttcggccaggccccgtc (SEQ ID NO: 571)	gttcggccaggccccgtcgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 592)
9	agtccccccccccccgc (SEQ ID NO: 572)	agtccccccccccccgcgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 593)
10	gtggggcactgaccgc (SEQ ID NO: 573)	gtggggcactgaccgcacgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCC (SEQ ID NO: 594)

11	aacctctaaggcaaaccttc (SEQ ID NO: 574)	aacctctaaggcaaacctcgttTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 595)
12	gttacacagaaccagaaggc (SEQ ID NO: 575)	gttacacagaaccagaaggcgTTAGAGCTAGAAATAGCAAGTTAAAATAA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTCC (SEQ ID NO: 596)
13	gaaggttcacacagaaccaga (SEQ ID NO: 576)	gaaggttcacacagaaccagatTTAGAGCTAGAAATAGCAAGTTAAAATAA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTCC (SEQ ID NO: 597)
14	agtcatgatgagtcatgtcg (SEQ ID NO: 577)	agtcatgatgagtcatgtcgTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 598)
15	gatgagtcatgtcgaggcTT (SEQ ID NO: 578)	gatgagtcatgtcgaggcTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 599)
16	actctaggctgagaacatct (SEQ ID NO: 579)	actctaggctgagaacatctcgTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 600)
17	gtccccagcaggatgttac (SEQ ID NO: 580)	gtccccagcaggatgttacggTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 601)
18	gccctgttaagcatactgtcg (SEQ ID NO: 581)	gccctgttaagcatactgtcgTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG GCTTTTTTCC (SEQ ID NO: 602)
19	cagggcagatggcaaaaaaaa (SEQ ID NO: 582)	cagggcagatggcaaaaaaaagtTTAGAGCTAGAAATAGCAAGTTAAAATAA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTCC (SEQ ID NO: 603)
20	gagggtggaggtttagtcagg (SEQ ID NO: 583)	gagggtggaggtttagtcaggTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTCC (SEQ ID NO: 604)
21	aaacggcatcataaagaaaaa (SEQ ID NO: 584)	aaacggcatcataaagaaaaagtTTAGAGCTAGAAATAGCAAGTTAAAATAA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTCC (SEQ ID NO: 605)

Detección de ARNg individuales en el locus de globina mediante CRISPR/dCas9. Los efectores dCas9 y dCas9-KRAB se suministraron lentivíricamente 5-8 días antes de la electroporación con 5 µg de un plásmido que codifica la expresión de U6-ARNg (Fig. 55A). Se incluyeron como controles las células que no se sometieron a electroporación

5 con ARNg (sin ARNg) y las células tratadas con un ARNg dirigido a un locus diferente (IL1RN). Múltiples ARNg efectuaron una potente represión de los genes de la ε-, γ- y β-globina cuando se ensayaron 3 días después de la transfección, con hasta un 80 % de atenuación lograda (Fig. 55B, 55C, 55D). La expresión de un ARNg con dCas9 o dCas9-KRAB inhibió la expresión génica en el locus de globina. Generalmente, el tratamiento con dCas9-KRAB dio como resultado una represión más fuerte para un ARNg determinado en comparación con dCas9 solo, lo que sugiere

10 un papel importante para el dominio KRAB en el reclutamiento de factores de heterocromatina que mejoran la represión. Los niveles de represión alcanzados dependen de la cantidad de plásmido de ARNg suministrado mediante transfección sólo en las células tratadas con dCas9-KRAB (Fig. 56A, 56B, 56C). Aumentar la dosis de plásmido Cr4 ARNg hasta 10 µg aumenta los niveles de silenciamiento de los genes de globina en células tratadas con dCas9-KRAB.

15 **Silenciamiento estable de genes de globina por dCas9-KRAB.** dCas9/dCas9-KRAB se coexpresaron con ARNg individuales de forma lentivírica en K562 (Fig. 57A). Se incluyeron como controles células que no fueron tratadas con lentivirus (NT), tratadas con dCas9/dCas9-KRAB sin ARNg (sin ARNg) y con dCas9/dCas9-KRAB y ARNg dirigido a un locus diferente (IL1RN). Las células tratadas con lentivirus se seleccionaron del día 4 al 7. Múltiples ARNg efectuaron una potente represión transcripcional de los genes de la ε-, γ- y β-globina cuando se ensayaron 7 días después de la transducción, con hasta un 95 % de atenuación lograda (Fig. 57B, 57C, 57D). La expresión de ε-globina fue la que más se silenció en respuesta a los ARNg dirigidos al potenciador HS2. El tratamiento con dCas9-KRAB con un ARNg resultó en una represión drásticamente mayor que el tratamiento con dCas9 y ARNg.

25 Estos descubrimientos demuestran que dCas9-KRAB dirigido al potenciador HS2 mediante ARNg produce una potente

represión de los genes de globina distal. Este es el primer ejemplo de control epigenético dirigido de elementos reguladores distales en células de mamíferos mediante el sistema CRISPR/Cas9. Los potenciadores regulan el desarrollo y la enfermedad, y esta divulgación proporciona un método para sondar y controlar la función potenciadora y puede usarse para determinar los efectos de dCas9-KRAB sobre la accesibilidad local a la cromatina y la expresión en todo el genoma.

Ejemplo 23

dCas9-p300

Se diseñó una proteína de fusión dCas9-p300 y se comparó con la proteína de fusión dCas9-VP64 (véase la Fig. 59). Las construcciones de aminoácidos de las construcciones dCas9 se muestran en las Fig. 61A-61C. Las células de la línea celular HEK293T de tejido de riñón embrionario humano (*Human Embryonic Kidney*) (ATCC; CRL-11268) se sembraron a una densidad de 1,5e5 células por pocillo en placas de cultivo de tejidos de 24 pocillos un día antes de la transfección con el reactivo de transfección Lipofectamine 2000 (Life Technologies). 24 h más tarde las células se transfecaron con 1 µl de Lipofectamine 2000, 375 ng de construcción de expresión de dCas9 (dCas9, dCas9VP64 o dCas9p300 respectivamente) y 125 ng de plásmidos de expresión de ARNg agrupados (4 de cada uno en proporciones equimolares). La Tabla 13 muestra la información de ARNg.

Tabla 13 Información de ARNg

<u>Ubicación de la diana</u>	<u>Secuencia del protoespaciador (5' - 3')</u>	<u>Ubicación genómica (Ensamblaje primario GRCh38)</u>	<u>Referencia</u>
Promotor IL1RN	TGTACTCTCTGAGGTGCTC (SEQ ID NO: 606)	Cr2: 113117865-113117883	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor IL1RN	ACGCAGATAAGAACCAAGTT (SEQ ID NO: 607)	Cr2: 113117714-113117732	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor IL1RN	CATCAAGTCAGCCATCAGC (SEQ ID NO: 608)	Cr2: 113117781-113117799	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor IL1RN	GAGTCACCCTCCTGGAAAC (SEQ ID NO: 609)	Cr2: 113117749-113117767	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor de MyoD	CCTGGGCTCCGGGGCGTTT (SEQ ID NO: 610)	Cr11: 17719509-17719527	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor de MyoD	GGCCCCTGCGGCCACCCCG (SEQ ID NO: 611)	Cr11: 17719422-17719440	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor de MyoD	CTCCCTCCCTGCCCGTAG (SEQ ID NO: 612)	Cr11: 17719350-17719368	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor de MyoD	AGGTTGGAAAGGGCGTGC (SEQ ID NO: 613)	Cr11: 17719290-17719308	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Promotor Oct4	ACTCCACTGCACTCCAGTCT (SEQ ID NO: 614)	Cr6: 31170953-31170934	Hu <i>et al.</i> , Nucleic Acids Res, 2014
Promotor Oct4	TCTGTGGGGGACCTGCACTG (SEQ ID NO: 615)	Cr6: 31170885-31170866	Hu <i>et al.</i> , Nucleic Acids Res, 2014
Promotor Oct4	GGGGCGCCAGTTGTCTCC (SEQ ID NO: 616)	Cr6: 31170855-31170836	Hu <i>et al.</i> , Nucleic Acids Res, 2014
Promotor Oct4	ACACCATTGCCACCACCAT (SEQ ID NO: 617)	Cr6: 31170816-31170797	Hu <i>et al.</i> , Nucleic Acids Res, 2014

Las células 3 días después de la transfección se recogieron y se analizaron mediante RT-QPCR para determinar la expresión de ARNm. Las secuencias del cebador RT-QPCR se enumeran en la Tabla 14.

Tabla 14 Cebadores RT-QPCR

<u>Cebador diana</u>	<u>Secuencia del cebador (5' - 3')</u>	<u>Referencia</u>
----------------------	--	-------------------

GAPDH Directo	CAATGACCCCTTCATTGACC (SEQ ID NO: 618)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
GAPDH Inverso	TTGATTTGGAGGGATCTCG (SEQ ID NO: 619)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
IL1RN Directo	GGAATCCATGGAGGGAAAGAT (SEQ ID NO: 620)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
IL1RN Inverso	TGTTCTCGCTCAGGTCAGTG (SEQ ID NO: 621)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
MyoD Directo	CTCTCTGCTCCTTGCCACA (SEQ ID NO: 622)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
MyoD Inverso	GTGCTCTCGGGTTTCAGGA (SEQ ID NO: 623)	Perez-Pinera <i>et al.</i> , Nat Methods, 2013
Oct4 Directo	CGAAAGAGAAAGCGAACCAAGTATCGAGAAC (SEQ ID NO: 624)	Hu <i>et al.</i> , Nucleic Acids Res, 2014
Oct4 Inverso	CGTTGTGCATACTCGCTGCTTGATCGC (SEQ ID NO: 625)	Hu <i>et al.</i> , Nucleic Acids Res, 2014

RT-QPCR se normalizó a la expresión de *GAPDH* usando el método $\Delta\Delta C_t$. Los resultados se expresan como factor de multiplicación de la expresión del gen de interés con respecto a las células tratadas con Lipofectamine solo sin ADN transfectado ("Sin ADN") (véanse las Fig. 60A-60C).

5 La Fig. 62 muestra que los residuos mutantes en el dominio HAT de p300 provocan una pérdida de su capacidad para activar la expresión génica. La Fig. 63 muestra que múltiples ARNg funcionan sinérgicamente con dCas-9-p300, como se muestra con dCas-9-VP64.

10 Ejemplo 24

La Fig. 66 muestra la integración mediada por TALEN de minidistrofina en la 5'UTR de la isoforma de distrofina del músculo esquelético Dp427m en líneas celulares de mioblastos esqueléticos derivadas de pacientes humanos con DMD que portan diferentes eliminaciones en el gen de la distrofina. Las células de pacientes con DMD se sometieron a electroporación con construcciones que codifican un par TALEN activo en el locus 5'UTR y una plantilla donante que porta el gen de minidistrofina. La Fig. 66(a) muestra un esquema de cómo se integró la minidistrofina en la 5'UTR. La Fig. 66(b) muestra que las líneas celulares clonales resistentes a la higromicina se aislaron y examinaron mediante PCR para integraciones exitosas específicas de sitio en la 5'UTR usando los cebadores que se muestran en la Fig. 66(a). Los asteriscos indican clones seleccionados para análisis adicionales en la Fig. 66(c). La Fig. 66(c) muestra que los mioblastos DMD clonalmente aislados con eventos de integración detectados se diferenciaron durante 6 días y se evaluó la expresión de una etiqueta HA fusionada al extremo C terminal de minidistrofina.

25 Ha de entenderse que la descripción detallada anterior y los ejemplos adjuntos son meramente ilustrativos y no han de interpretarse como limitaciones en el alcance de la invención, que se define únicamente por las reivindicaciones adjuntas y sus equivalentes.

Apéndice

30 Secuencia de ADN hCas9-T2A-GFP: SEQ ID NO: 145 (Secuencia optimizada humana SpCas9, etiqueta HA, péptido T2A, secuencia eGFP)

gccaccATGGACAAGAAGTACTCCATTGGGCTCGATATCGCACAAACAGCGTCGGCTGGGCCGTATTACGGACGAGTA
 CAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGCAATACCGATGCCACAGCATAAGAAGAACCTCATGGGCCCTCC
 TGTTCGACTCCGGGGAGACGCCAAGGCCACGCCGCTCAAAAGAACAGCACGCCAGATATAACCGCAGAAAAGAATCGG
 ATCTGCTACCTGCAGGAGATCTTAGTAATGAGATGGCTAAGGTGGATGACTCTTCTTCCATAGGCTGGAGGAGTCCTT
 TTTGGTGGAGGAGGATAAAAAGCACGAGGCCACCAACTTTGGCAATATCGGACGAGGTGGCTACCATGAAAAGT
 ACCAACCATATATCATCTGAGGAAGAAGCTTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCCTG
 GCGCATATGATCAAATTCTGGGACACTTCTCATCGAGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTT
 TATCCAACCTGGTCAGACTTACAATCAGCTTCTGAAGAGAACCCGATCAACGATCCGGAGTTGACGCCAAAGCAATCC
 TGAGCGCTAGGCTGTCATACTCCGGCGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTT
 GGTAATCTTATGCCCTGTCACTCGGCTGACCCCCAACCTTAAATCTAACTTCGACCTGGCGAAGATGCCAAGCTTCA
 ACTGAGCAAAGACACCTACGATGATCTGACAATCTGCTGGCCAGATCGGCGACCAGTACGCAGACCTTTTGG
 CGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAAACCGGAGATACCAAAGCTCCGCTGAGC
 GCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTGACTTGCTGAAGGCCCTTGCTAGACAGCAACTGCCGTA
 GAAGTACAAGGAAATTCTTCGATCAGTCTAAATGGCTACGCCGATACATTGACGCCGAGCAAGCAGGAGGAAT
 TTTACAAATTATTAAAGCCATCTGGAAAAATGGACGGCACCGAGGAGCTGTTAAAGCTTAACAGAGAAGATCTG
 TTGCGAAACAGCGACTTTCGACAATGGAAGCATCCCCCACCAGATTACCTGGCGAAGTGCACGCTATCCTCAGGCG
 GCAAGAGGATTCTACCCCTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTGCGATACCCACTATG
 TAGGCCCTCGCCGGGAAATTCCAGATTGCGTGGATGACTCGAAATCAGAAGAGACCATCACTCCCTGGAAACTTC
 GAGGAAGTGTGGATAAGGGGCTCTGCCAGTCCTCATCGAAAGGATGACTAACTTGTATAAAATCTGCTAACAGA
 AAAGGTGTTCTAAACACTCTGCTGTACGAGTACTCACAGTTATAACGAGCTACCAAGGTCAAATACGTACAG
 AAGGGATGAGAAAGCCGACTTCTGTCTGGAGAGCAGAAGAAAGCTATGTCAGCTCCCTTCAAGACGAACCGGAA
 GTTACCGTAAACAGCTAAAGAAGACTATTCTAAAGATTGAATGTTGACTCTGTTGAAATCAGGGAGTGGAGGA
 TCGCTAACGCATCCGGAACGTACGATCTCTGAAATCATTAAAGACAAGGACTTCTGGACAATGAGGAGA
 ACGAGGACATTCTGAGGACATTGTCTCACCTTACGTTGAGATAGGGAGATGATTGAAAGAACGCTTGAAGAAA
 TACGCTCATCTCTGACGACAAAGTCATGAAACAGCTAACAGGGCCGATATAACAGGATGGGGCGGCTGTCAGAAA
 ACTGATCAATGGGATCCGAGACAAGCAGTGGAAAGACAATCTGGATTCTTAAAGTCCGATGGATTGCAACCCGGA
 ACTTCATGCACTGTTGATCCATGATGACTCTCACCCTTAAGGAGGACATCCAGAAAGCACAAGTCTGGCCAGGGGAC
 AGTCTTCAGGACACATCGCTAATCTGCAAGGTAGCCAGCTATCAAAGGAAACTCGCAGACCGTTAAGGTGTTG
 TGAACCTGCTAAAGTAATGGGAAGGCATAAGCCCAGAAATATCGTTATCGAGATGGCCGAGAGAACCAAACCTACCCAGA
 AGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAGAACCTGGGCTCCAAATCTTAAGGAA
 CACCCAGTGGAAAACACCCAGCTCAGATGAGGAGGTTGACTACCTGAGACAGGAGCATGTCAGTGG
 TCAGGAACCTGGACATCACTGGCTCTCGGACTACGACGGTGGATCATCTGTCAGGCTTCTTCTCAAAGATGATTCTA
 TTGATAATAAAAGTGTGACAAGATCCGATAAAAATAGAGGGAAAGAGTGATAACGTCCTCTGAGAAGGTTGTCAGAAA
 ATGAAAATATTGGCGGAGCTGCTGAACGCCAACTGATCACACACCGGAAGTCTGATAATCTGACTAAGGCTGAACG
 AGGTGGCCTGTCTGGATTAAGCCGCTTCATCAAAGGAGCTTGTGAGACACGCCAGATCACCAAGCACGTGG
 CCCAAATTCTCGATTACGACGAAACACCAAGTACGATGAAAATGACAAACTGATTGAGAGGTTGAAAGTTTACTCTG
 AAGTCTAAGCTGGTCTCAGATTCAGAAAGACTTTCAGTTTATAAGGTGAGAGAGATCAACAATTACCCATGCGCA
 TGATGCCCTACCTGAATGCAGTGGTAGGCAGTCAGTCTTACAAAAATATCCAGCTGAAATTGTTACCGGAG
 ACTATAAAGTGTACGATGTTAGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTCTT
 TACAGCAATATTATGAAATTCTCAAGACCGAGATTACACTGGCCAATGGAGAGATTGGAGAGCGACCACTTATCGAAC
 AAACGGGAAACAGGAGAAATCGTGTGGACAAGGGTAGGGATTCTCGAGCTCCGGAAAGTCTGTCCATGCCGAGG
 TGAACATCGTAAAGAACCGGAAGTACAGACCGGGAGGCTCTCCAAGGAAAGTATCTCCGAAAGAACAGCGACAAG
 CTGATCGCACGCAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCCTACAGTCGCTTACAGTGTACTGG
 TGTGGCCAAGTGGAGAAAGGAAAGTCTAAAAAAACTCAAAGCGTCAAGGAACCTGCTGGCATCACATCGAGGCGAT
 CAAGCTCGAAAAAAACCCCATCGACTTCTCGAGGCGAAAGGATATAAGGGTCAAAGAACCTCATCTTAAAGCTT
 CCCAAGTACTCTCTTCTGAGCTTGGAAACGGCCGAAACGAATGTCGCTAGTGCAGGCGAGCTGCAAGAACGGTAA
 GCTGGCACTGCCCTCAAATACGTTAATTCTTGATCTGGCAGCCACTATGAAAAGCTCAAAGGGTCTCCGAAAGATA
 ATGAGCAGAAGCAGCTGTTCTGGAAACAACACAAACTACCTTGATGAGATCATCGAGCAAATAAGCAATTCTCCAA
 AGAGTGTCTCGCCGACGCTAACCTCGATAAGGTGTTCTGCTTACAATAAGCACAGGGATAAGCCCATCGGGAGCA
 GGCAGAAAACATTATCCACTTGTGTTACTCTGACCAACTGGGCGCCCTGCGAGCTTCAAGTACTTCGACACCACCATAG
 ACAGAAAGCGGTACACCTCTACAAAGGAGGTCTGGACGCCACACTGATTGATCATAGTCATTACGGGCTCTATGAAACA
 AGAATCGACCTCTCTAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGGTAGCGAGGGCAGAG
 AAGTCTCTAACATCGGGTGAGTGGAGGAATCCGCCCTGgtacCgtgagcaaggcgagggactgtttcacgggg
 tggtgcccatccctggcgcagctggacggcgcacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgcc
 acctacggcaagctgaccctgaagttcatctgcaccacccggcaagctgcccgtccctggccaccctgtgaccaccc
 gacctacggcgtgcagtgcattcagccgtaccccgaccacatgaagcagcacgacttctcaagtcgccccatgcccc
 gctacgtccaggagcgcacccatcttcaaggacgcacggcaactacaagacccgcgcggaggtgaagttcgagggcgac
 accctggtaaccgcacgtgagctgaaaggcatcgacttcaaggaggacggcaacatccctggggcacaactgtggagtt
 ctacaacagccacaacgtctatcatggccgacaaggagaacacggcatcaaggtaacttcaagatccgccccacaaca
 tcgaggacggcagcgtgcagctccggccgaccactaccacgcagaacaccccatggcgacggccccgtgtgtcgcccc
 aaccactacgtggacccacttccgcctgagcaaagaccccaacggagaagcgcgcacatggcctgtgtggagtt
 gaccggccggatcactctcggcatggacgcgtgtacaagAccggTTAG

ES 2 987 399 T3

ES 2 987 399 T3

Secuencia diana del potenciador HS2 (SEQ ID NO:467)

taagcttcagttttcccttagttcctgttacatttctgtgtgtctccattagtgacccatccatgtccaaggcatgagcagtctggccaggccccctgtcggggtcagtgccccaccccccgccttctggttctgtgtaaccttctaagcaaaccttctggctcaagcacagcaatgctgagtcatgatgagtcatgctgaggcttaggggtgtgtgcccagatgttctcagcctagagtatgatgactcctatctgggtccccagcaggatgctacagggcagatggaaaaaaaaaggagaagctgaccacctgactaaaactccacctcaaacggcatataaaagaaaaatggatgcctgagacagaatgtgacatattctagaatatatt

5

Secuencia dSpCas9-KRAB (SEQ ID NO:466)

ATGGACTACAAAGACCATGACGGTATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAtggccccaa
 gaagaagaggaaggtggccgcggaaATGGACAAGAAGTACTCCATTGGCTGCCATGGCACAAACAGCGTCGGCTGGG
 CCGTCATTACGGACGGTACAAGGTGCCAGCAGAAAAATTCAAAGTTCTGGCAATACCGATGCCACAGCATAAAAGAAG
 AACCTCATTGGGCCCTCCTGTCGACTCCGGGAAACCGCGAAGCCACGCCGCTCAAAGAAGAACAGCACGGCGAGATA
 TACCCGAGAAAGAATCGGATCTGCTACCTcgAGAGATCTTAGTAATGAGATGGCTAAGGTGGATGACTCTTCTCC
 ATAGGCTGGAGGAGTCCTTTGGTGGAGGAGATAAAAGCAGCGGCCACCAATCTTGGCAATATCGTGGACGAG
 GTGGCGTACCATGAAAAGTACCCAACCATAATCATCTGAGGAAGAAGCTTGAGACAGTACTGATAAGGCTGACTGCG
 GTTGTATCTCGCGCTGGCGATATGATCAAATTGCGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACA
 GCGATGTCGACAAACTCTTATCCAAGTGGTCAGACTTACAATCAGCTTCTGAAGAGAACCGATCAACGCATCCGA
 GTTGACGCCAAGCAATCCTGAGCGCTAGGCTGTCAAATCCGGCGCTCGAAAACCTCATCGCACAGCTCCCTGGGA
 GAAGAAGAACGGCTGTTGGTAATCTTATGCCCTGTCAGTCGGCTGACCCCCACTTAAATCTAACTTCGACCTGG
 CGAAGATGCCAAGCTCAACTGAGCAAAGACACCTACGATGATGATCTGACAATCTGCTGGCCAGATCGCGACCAAG
 TAGCAGACCTTTTGGCGGAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGACACCGAGAT
 CACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACAGACTGACTTTGCTGAAGGCCCTG
 TCAGACAGCAACTGCCGTAGAAGTACAAGGAAATTTCCTCGATCAGTCTAAAGGCTACGCCGGATACATTGACGGC
 GGAGCAAGCCAGGAGGAATTTCACAAATTATAAGCCATCTTGAAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAA
 GCTTAACAGAGAAGATCTGTTGCGAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGCGAAC
 TGCACGCTATCCTCAGCGGCAAGAGGATTCTACCCCTTTGAAAGATAACAGGAAAAGATTGAGAAAATCCTCACA
 TTTCGGATACCCACTATGTAGGCCCTCGCCCAGGGAAATTCCAGATTGCGTGGATGACTCGCAAATCAGAAGAGAC
 CATCACTCCCTGGAACTTCGAGGAAGTCGTTGATAAGGGGCTCTGCCAGTCCTCATCGAAAGGATGACTAACTTTG
 ATAAAAAATCTGCTAACGAAAGGTGCTCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTATAACGAGCTCACC
 AAGGTCAAATACGTACAGAAGGGATGAGAAAGCCAGATTCTGTCTGGAGAGCAGAAGAAAGCTATCGGACCTCCT
 CTTCAAGACGAACGGGAAAGTTACCGTGAACAGCTCAAAGAAGACTATTTCAAAAGATTGAATGTTGACTCTGTTG
 AAATCAGCGGAGTGGAGGATCGCTAACGCATCCCTGGAAAGTACGATCAGTCTGAAAGGATGACTAACTTTG
 TTCCTGGACAATGAGGAGAACGAGGACATTCTGAGGACATTGCTCACCCTACGTTGTTGAAGATAGGGAGATGAT
 TGAAGAACGCTTAAAAGCTTACGCTCATCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCCGATATACAGGAT
 GGGGGCGCTGTCAGAAAAGTGTCAATGGatcCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTCCTTAAGTCC
 GATGGATTGCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTAAGGAGGACATCCAGAAAGCACA
 AGTTTCTGCCAGGGGGACAGTCTTCAGCAGCACATCGTAATCTGAGGTAGCCAGCTATCAAAGGGAATACTGC
 AGACCGTTAAGGTGGATGAACTCGTCAAAGTAATGGGAAGGATAAGCCCGAGAATATCGTTATCGAGATGGCCGA
 GAGAACCAAACCTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAAC
 ACTGGGTCTCCAAATCTTAAGGAAACCCAGTTGAAACACCCAGCTCAGAATGAGAACAGCTCTACTGCGAACAG
 GCAGGGACATGTACGTGGATCAGGAACCTGGACATCAATCGCTCCGACTACGACGTGGATGCCATCGTGCCTCCAGTCT
 TTTCTCAAAGATGATTCTATTGATAATAAAAGTGTGACAAGATCGGATAAAAATAGAGGGAAAGAGTGATAACGTC
 AGAAGAACGTTGTCAAGAAAATGAAAATTGGCGGAGCTGCTGAACGCCAAACTGATCACACAACGGAAAGTTGATA
 ATCTGACTAAGGCTGAACGGGGTGGCTGTCAGTTGAGATAAGGGCTTCATCAAAGGCAGCTTGTGAGACACCGC
 CAGATCACCAAGcaCGTGGCCAAATTCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTGAGA
 GGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTGAGAAAGACTTCAGTTTATAAGGTGAGAGAGATCA
 ACAATTACCCACCATGCGCATGATGCCCTACCTGAATGCGACTGGACTGCACTTATCAAAGGAAAGTCTGAA
 TCTGAATTGTTACGGAGACTATAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGG
 CACCGCTAAGTACTCTTTACAGCAATATTGAAATTGAAATTGAGGAAAGACCGAGATTACACTGCCAATGGAGAGATT
 AGCGACCACTTATGAAACAAACGGAGAACAGGAGAAATCGTGGAGACAAGGGTAGGGATTTCGCGACAGTCCGGAAG
 GTCCCTGTCATGCCAGGTGAACATCGTTAAAAGACCGAAGTACAGACCGGAGGCTTCTCAAGGAAAGTATCCTCCC
 GAAAAGAACAGCGACAAGCTGATCGCACGCAAAAAGATTGGACCCCAAGAAATACGGGGATTCTGATTCTCCTACAG
 TCGTTACAGTGTACTGGTTGGCCAAAGTGGAGAAAGGGAAGTCTAAAGGACTCAAAGCGTCAAGGAAGTCTGGGC
 ATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCATCGACTTCTCGAGGGGAAAGGATATAAGGAGTCAA
 AGACCTCATCATTAAGCTCCCAAGTACTCTCTTGTAGCTGAAAGGAAACGGCCGAAACGAATGCTCGTAGTGC
 AGCTGAGAACGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTCTGTATCTGGCAGGCCACTATGAAAAGCTC
 AAAGGGCTCCCGAAGATAATGAGCAGAAGCAGCTGTTGAGAACACAAACACTACCTTGATGAGATCATCGAGCA
 AATAAGCGAATTCTCCAAAAGAGTGTACCTCGCCGACGCTAACCTCGATAAGGTGTTCTGCTTACAATAAGCACAGGG
 ATAAGCCCATCAGGGAGCAGGAGAAAACATTATCCACTTGTGTTACTCTGACCAACTGGGGCGCCTGCGCCTCAAG
 TACTTCGACACCACCATAGACAGAACAGGGTACACCTCTACAAAGAGGTCTGGACGCCACACTGATTGATCAGTC
 ATACGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAC
 TGGctagCGATGCTAAGTCACTGACTGCCTGGTCCGGACACTGGTGACCTTCAAGGATGTGTTGTGGACTTCA
 CAGGAGTGGAAAGCTGCTGGACACTGCTCAGCAGATCTGTACAGAAATGTGATGCTGGAGAACTATAAGAAC
 CTTGGGTTATCAGCTTAAGCAGATGTGATCCTCCGGTTGGAGAAGGGAGAACGCCCTGGCTGGTGGAGAGAGAAA
 TTCACCAAGAGAGACCCATCCTGATTAGAGACTGCATTGAAATCATCAGTTCCGAAAAAGAACCGAAAGtGct
 agCG

REIVINDICACIONES

1. Un sistema de direccionamiento de ADN que se une a un gen de gamma globina humana, comprendiendo el sistema de direccionamiento de ADN una proteína de fusión Cas y al menos un ARN guía (ARNg) que se dirige a una secuencia polinucleotídica con SEQ ID NO: 33, 34, 35 o 36, en donde el gen de gamma globina es HBG1 y/o HBG2.
2. El sistema de direccionamiento de ADN de la reivindicación 1, en donde la proteína de fusión Cas comprende dos dominios polipeptídicos heterólogos, en donde el primer dominio polipeptídico comprende una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (Cas) y el segundo dominio polipeptídico tiene una actividad seleccionada del grupo que consiste en actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa y actividad desmetilasa.
3. Un sistema de direccionamiento de ADN que se une a un gen de gamma globina humana, el sistema de direccionamiento de ADN que comprende al menos un ARN guía (ARNg) y una proteína de fusión que comprende dos dominios polipeptídicos heterólogos, en donde el primer dominio polipeptídico comprende una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (Cas) y el segundo dominio polipeptídico tiene una actividad seleccionada del grupo que consiste en actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de asociación de ácidos nucleicos, actividad metilasa y actividad desmetilasa, en donde al menos un ARN guía (ARNg) se dirige a una secuencia polinucleotídica con SEQ ID NO: 33, 34, 35 o 36 y en donde el gen de gamma globina es HBG1 y/o HBG2.
4. El sistema de direccionamiento de ADN de la reivindicación 1 o 3, en donde el al menos un ARNg se dirige a una secuencia polinucleotídica con:
 - a) un complemento de SEQ ID NO: 33, 34, 35 o 36;
 - b) un ácido nucleico que es sustancialmente idéntico a SEQ ID NO: 33, 34, 35 o 36 o complemento de las mismas; o
 - c) un ácido nucleico que se hibrida en condiciones estrictas con SEQ ID NO: 33, 34, 35 o 36, complemento de las mismas o una secuencia sustancialmente idéntica a las mismas.
5. El sistema de direccionamiento de ADN de la reivindicación 2 o 3, en donde la proteína Cas comprende una Cas9, en donde la Cas9 comprende al menos una mutación de aminoácido que anula la actividad nucleasa de Cas9.
6. El sistema de direccionamiento de ADN de la reivindicación 5, en donde la al menos una mutación de aminoácido es al menos una de D10A y H840A.
7. El sistema de direccionamiento de ADN de la reivindicación 6, en donde la proteína Cas comprende iCas9 (aminoácidos 36-1403 de SEQ ID NO: 1).
8. El sistema de direccionamiento de ADN de la reivindicación 2 o 3, en donde el segundo dominio polipeptídico de la proteína de fusión comprende al menos una repetición del dominio de activación de la transcripción VP16 o un dominio KRAB.
9. El sistema de direccionamiento de ADN de la reivindicación 8, en donde el segundo dominio polipeptídico de la proteína de fusión comprende un tetrámero de VP16 ("VP64") o un dominio de activación de p65.
10. El sistema de direccionamiento de ADN de la reivindicación 2 o 3, en donde la proteína de fusión comprende además un enlazador que conecta el primer dominio polipeptídico con el segundo dominio polipeptídico.
11. El sistema de direccionamiento de ADN de la reivindicación 1 o 3, en donde el al menos un ARNg se dirige a una región promotora de HBG1 y/o HBG2.
12. Un polinucleótido aislado que codifica el sistema de direccionamiento de ADN de la reivindicación 1 o 3.
13. Un vector que comprende el polinucleótido aislado de la reivindicación 12.
14. Una célula que comprende el polinucleótido aislado de la reivindicación 12.
15. Un método para modular la expresión de genes de mamíferos en una célula, comprendiendo el método poner en contacto la célula con el sistema de direccionamiento de ADN de la reivindicación 1 o 3.
16. Una composición para inducir la expresión génica en un sujeto, comprendiendo la composición el sistema de direccionamiento de ADN de la reivindicación 1 o 3.
17. Un vector lentivírico modificado que comprende un polinucleótido aislado que codifica el sistema de

direcciónamiento de ADN de la reivindicación 2 o 3, comprendiendo el polinucleótido aislado una primera secuencia polinucleotídica que codifica la proteína de fusión y una segunda secuencia polinucleotídica que codifica el al menos un ARNg.

- 5 18. El vector lentivírico modificado de la reivindicación 17 para su uso en un método de activación de la expresión génica endógena en un sujeto.
19. El vector lentivírico modificado de la reivindicación 18 para el uso de la reivindicación 18, en donde el sujeto padece enfermedad de células falciformes o talasemia.

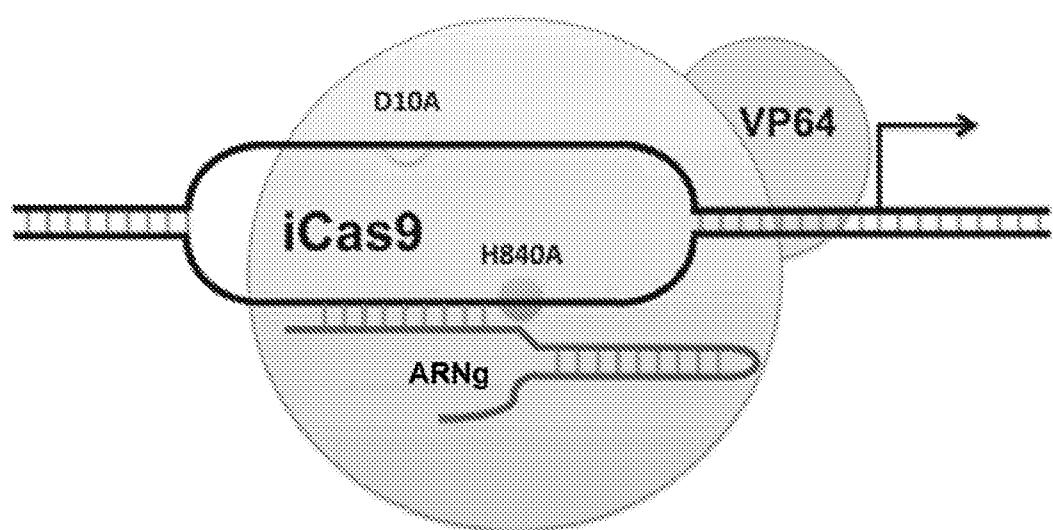


Fig. 1A

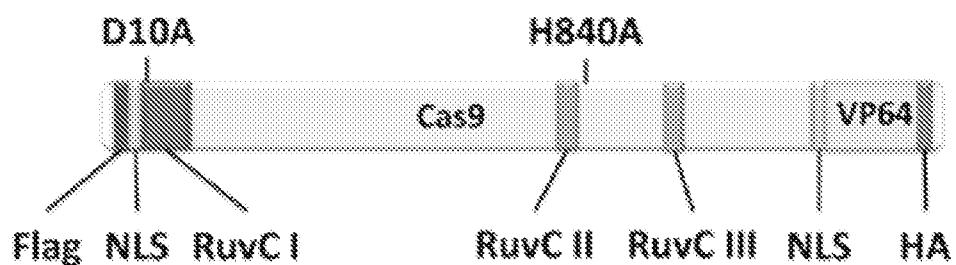
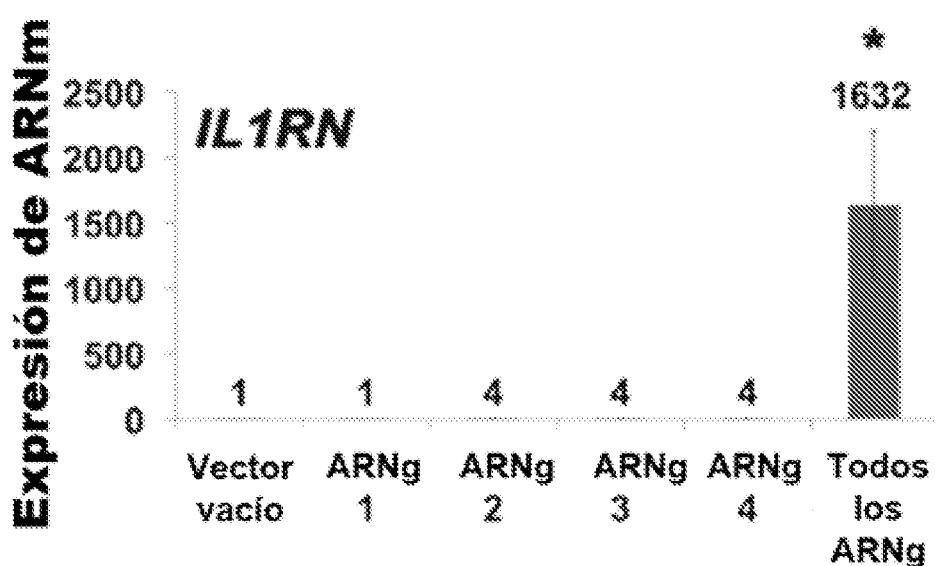
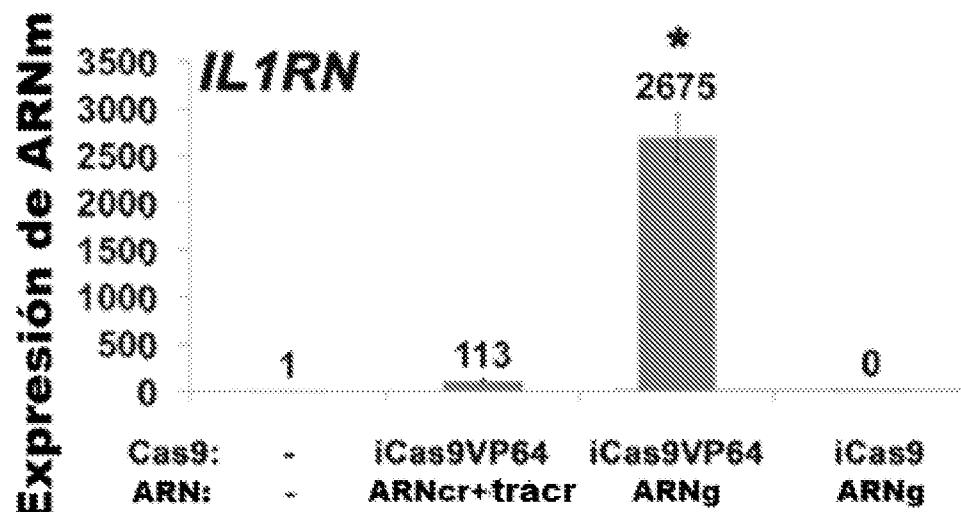


Fig. 1B



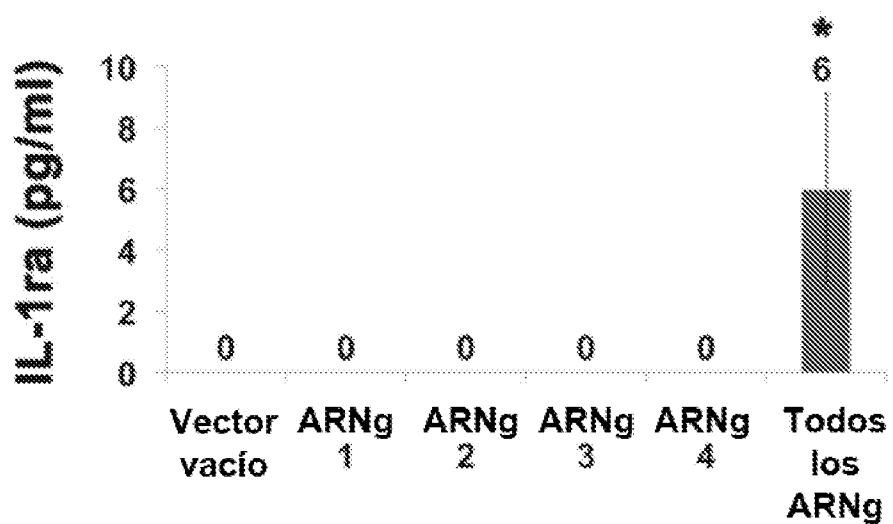


Fig. 1E

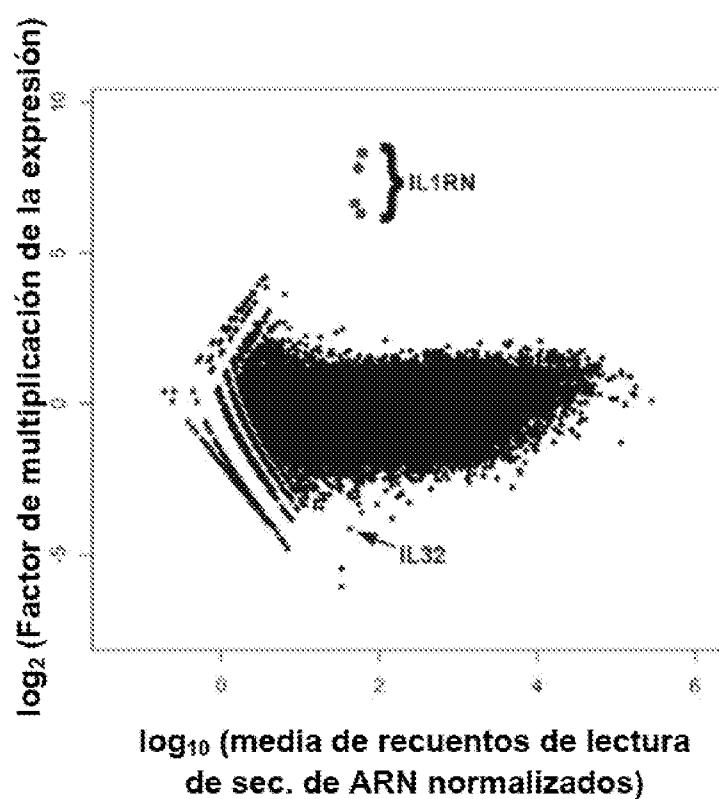


Fig. 1F

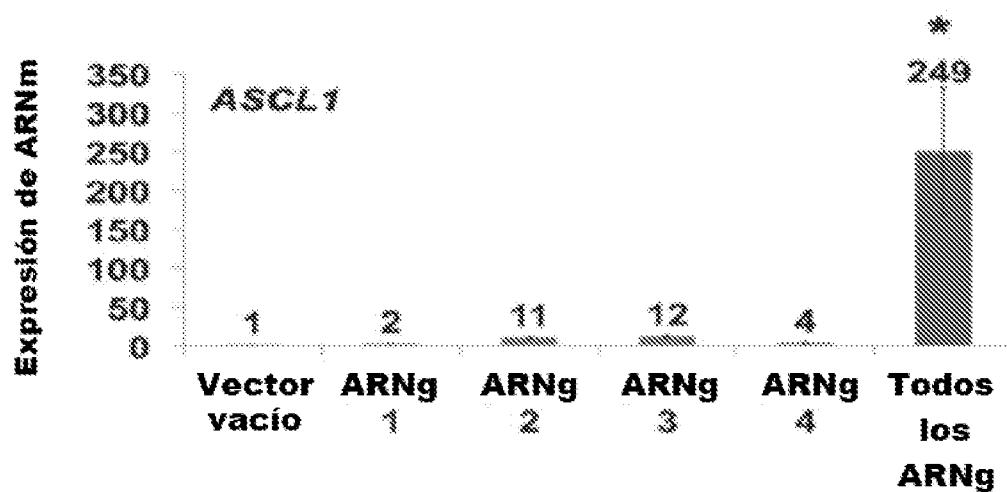


Fig. 2A

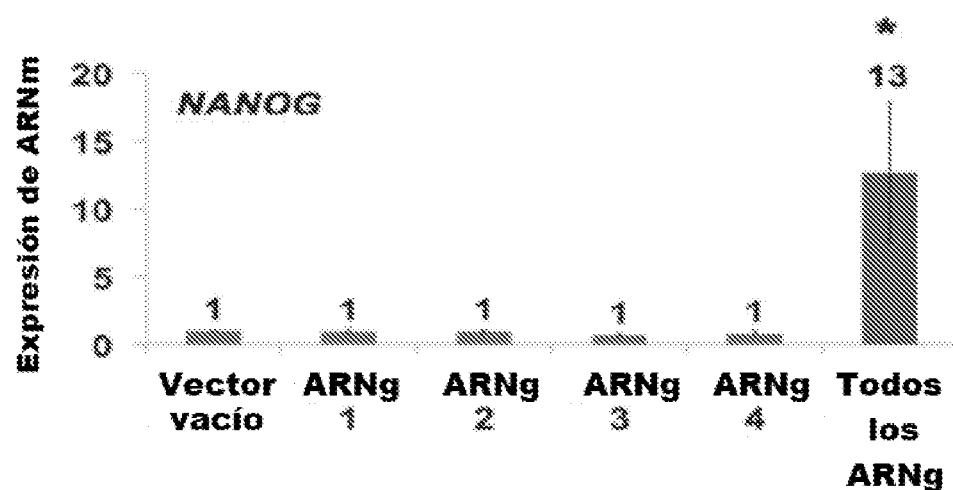
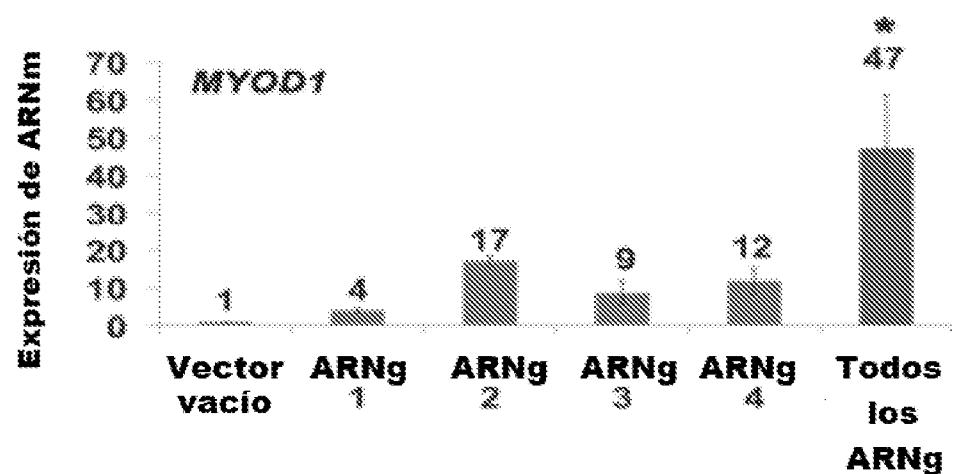
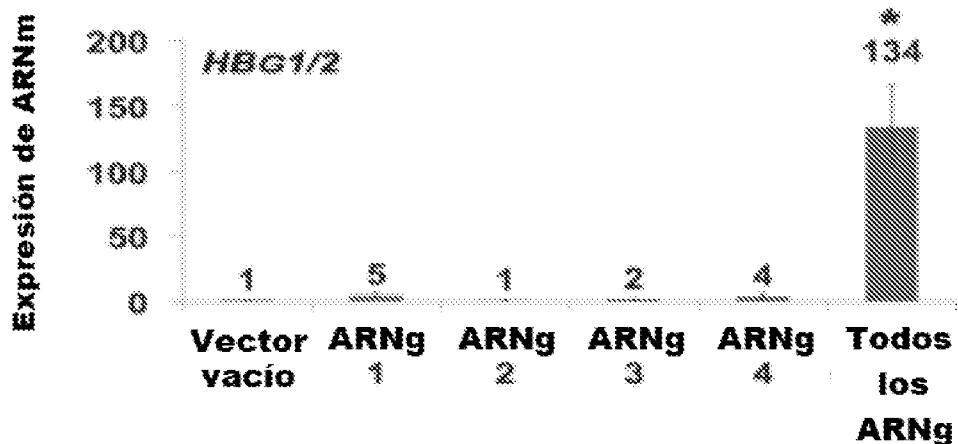
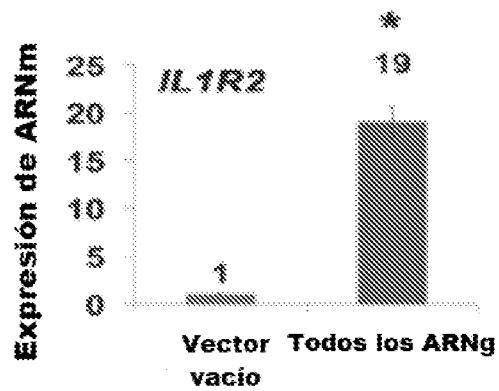
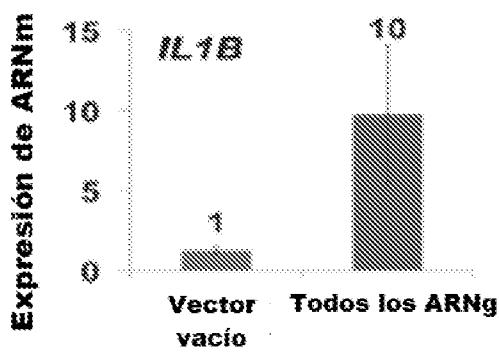
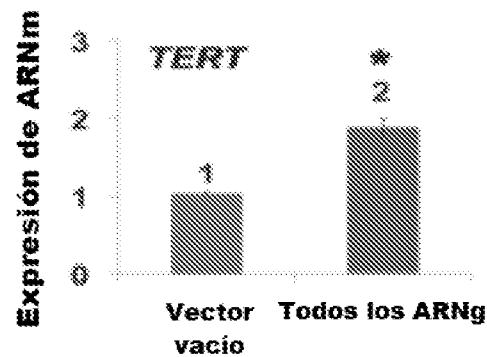
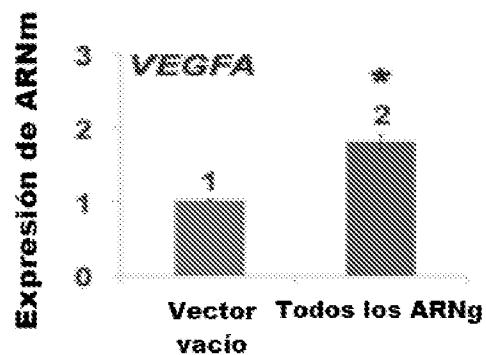


Fig. 2B





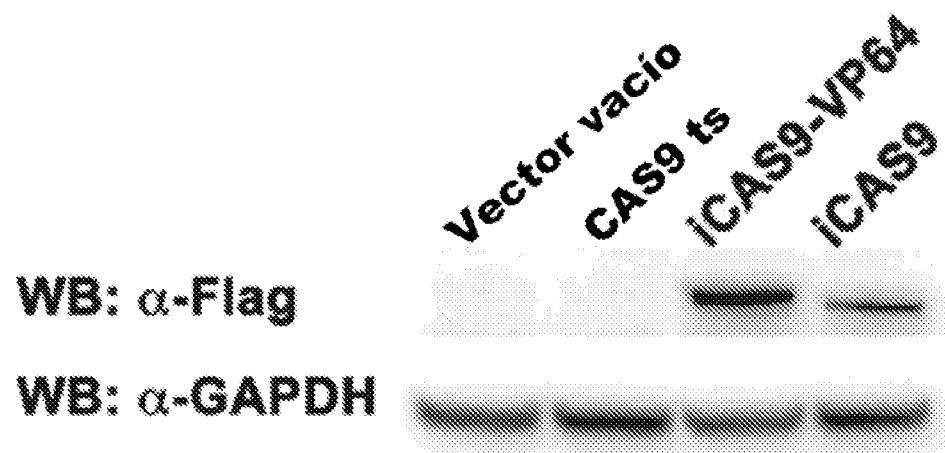
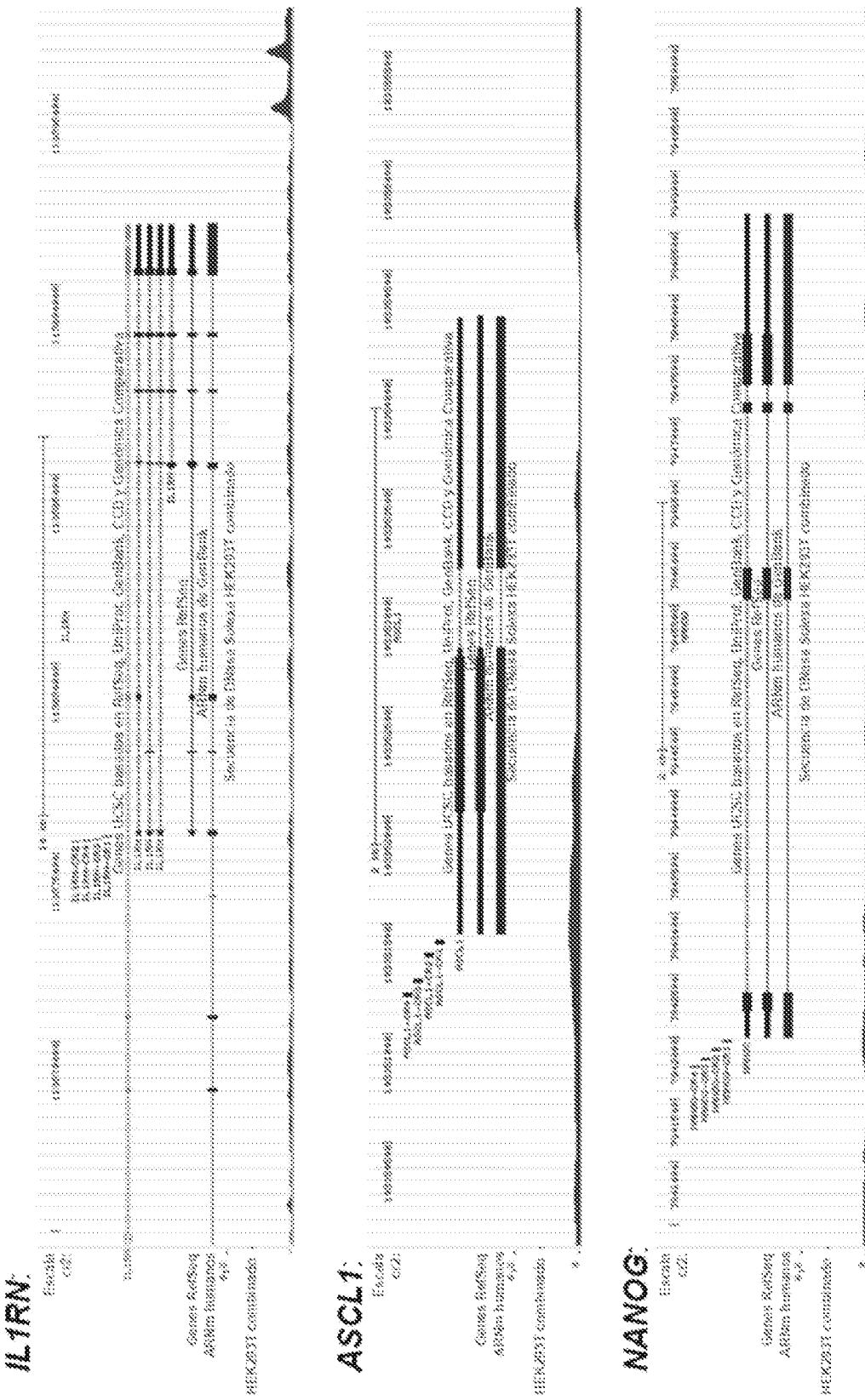


Fig. 3

ES 2 987 399 T3



卷之三

ES 2 987 399 T3

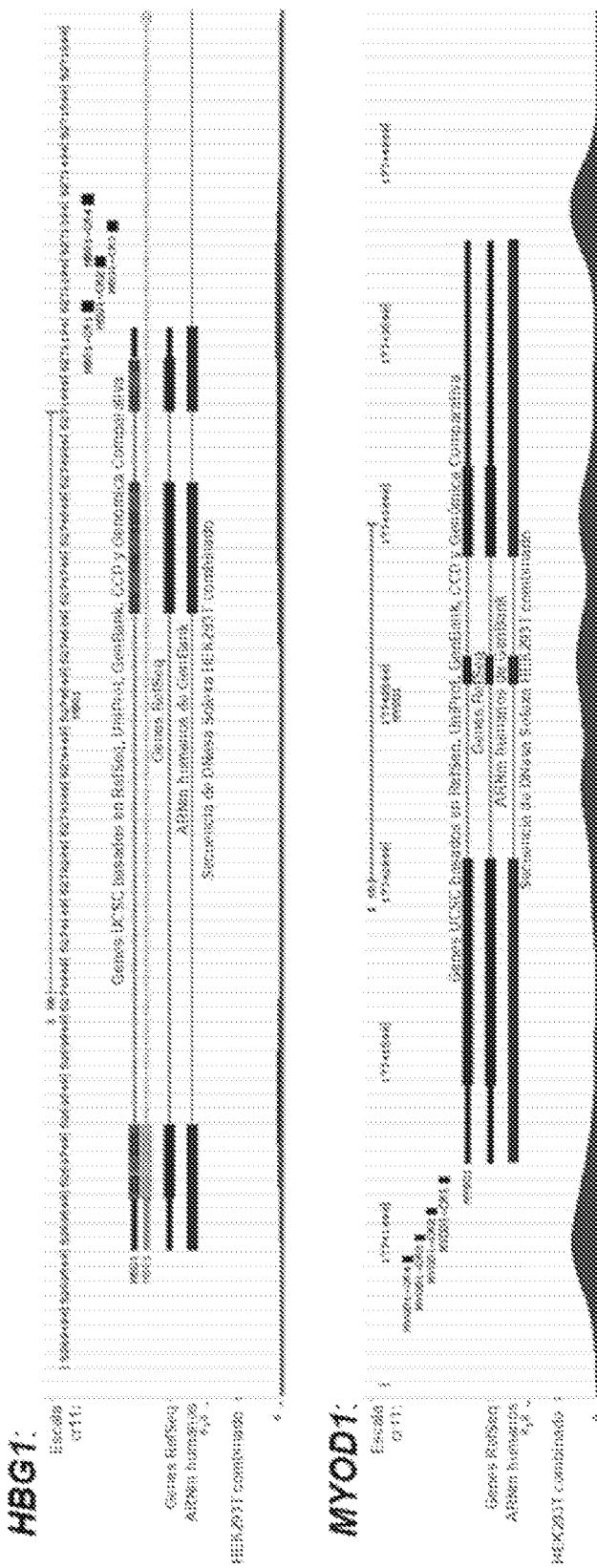
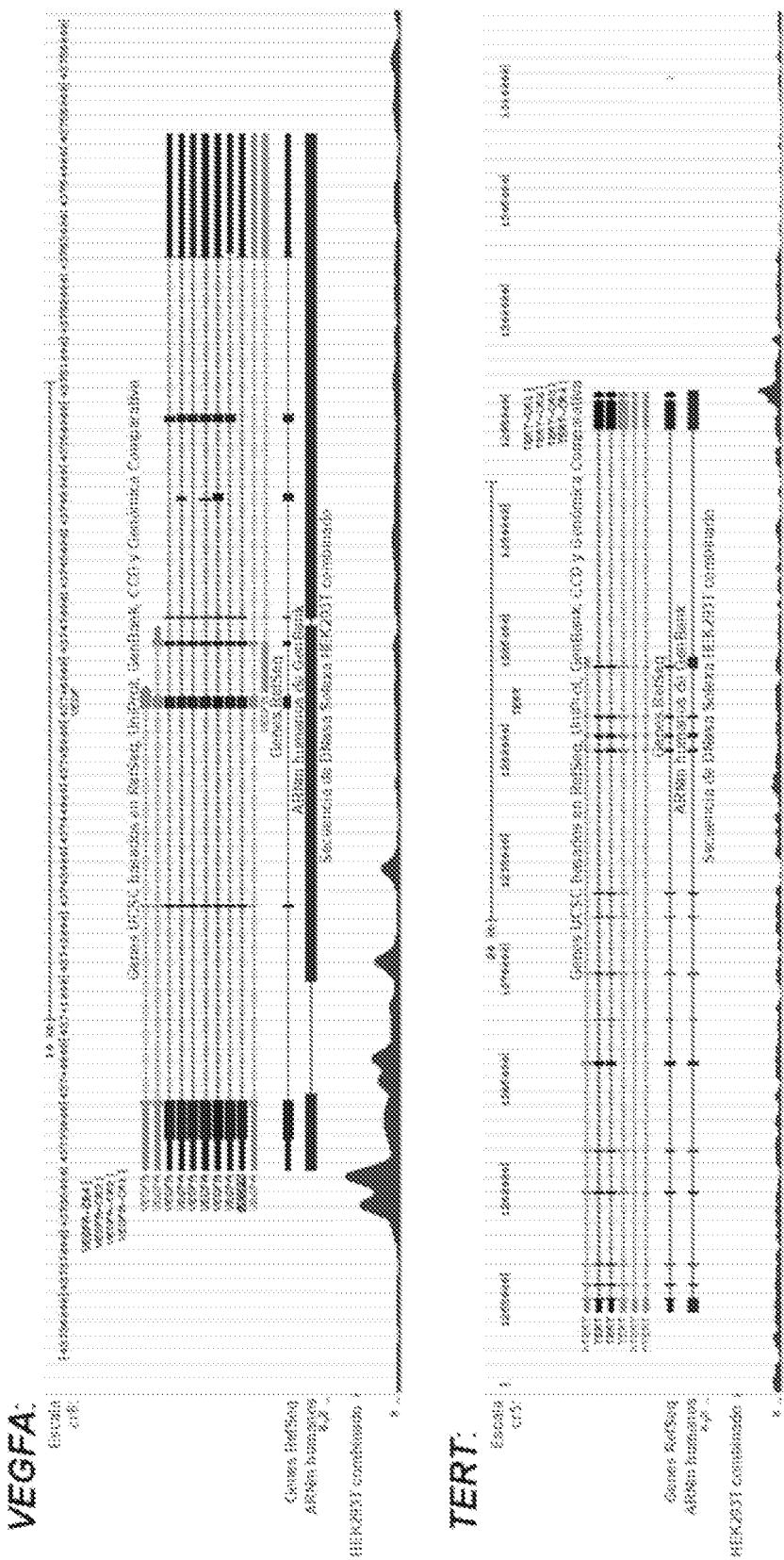


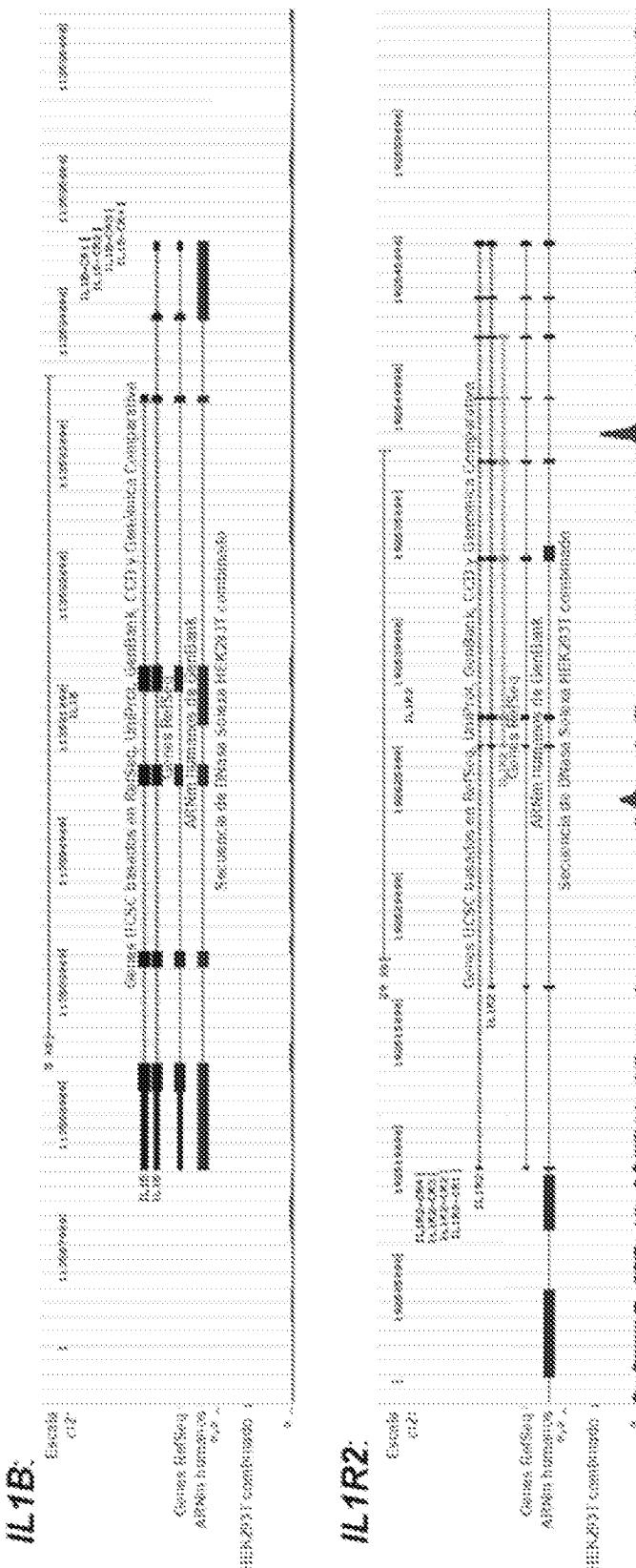
Fig. 43

ES 2 987 399 T3



四〇

ES 2 987 399 T3



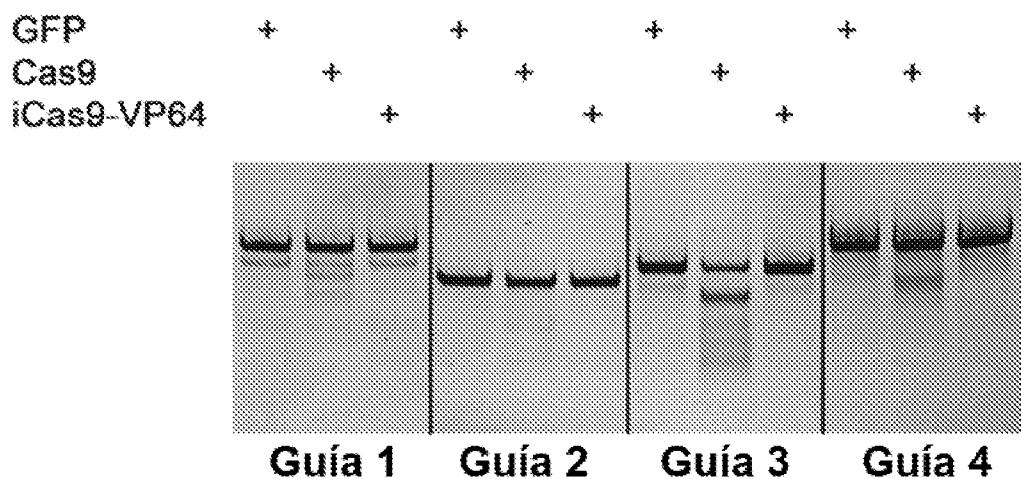


Fig. 5

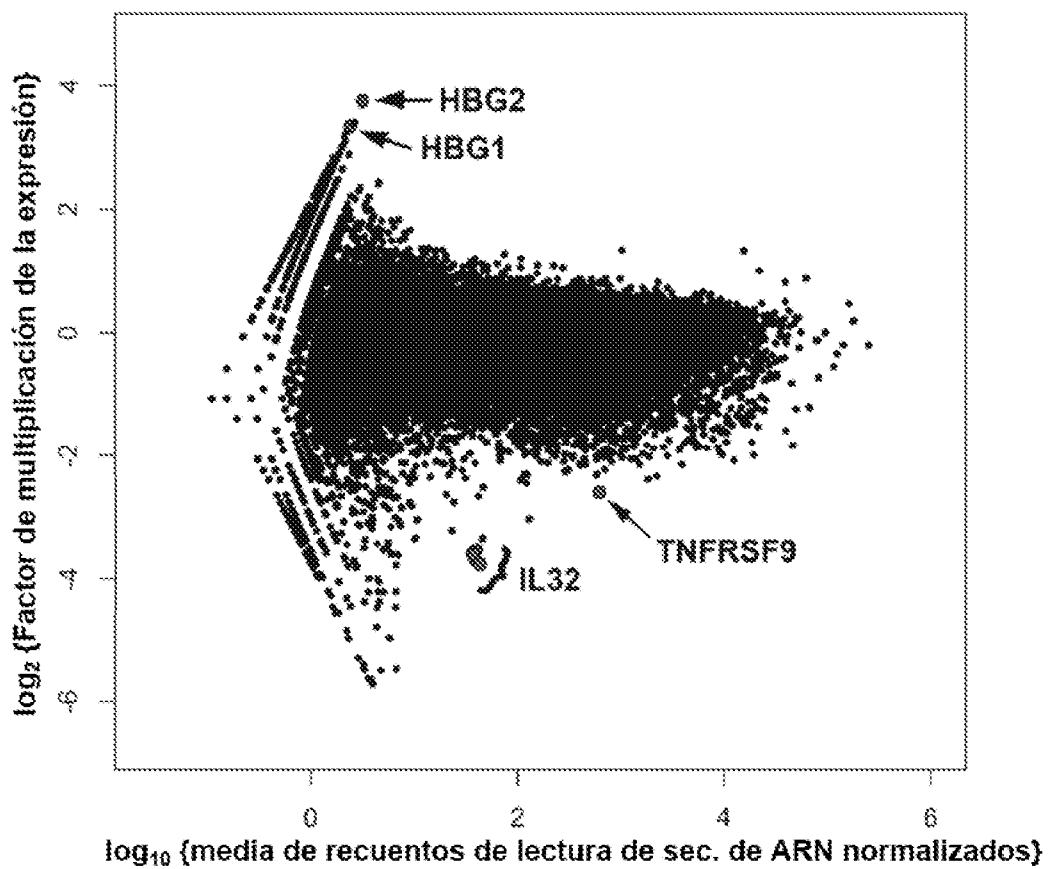


Fig. 6

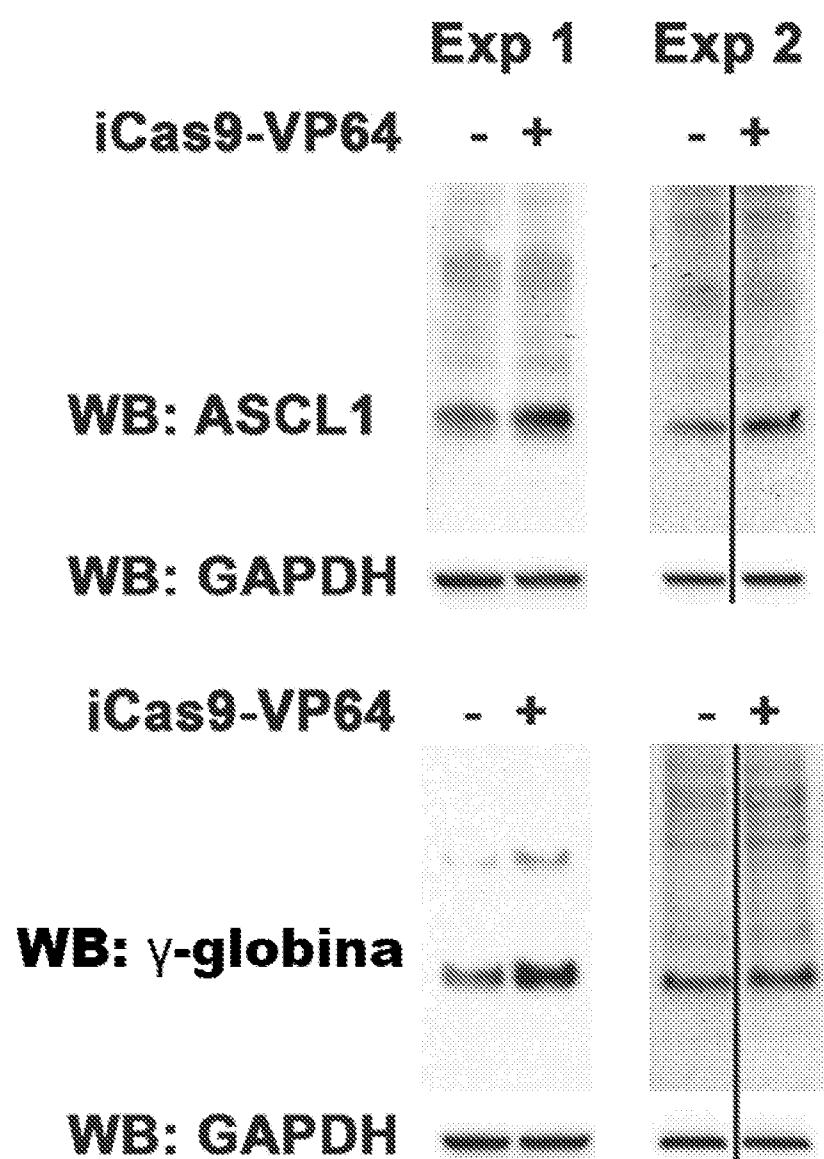


Fig. 7

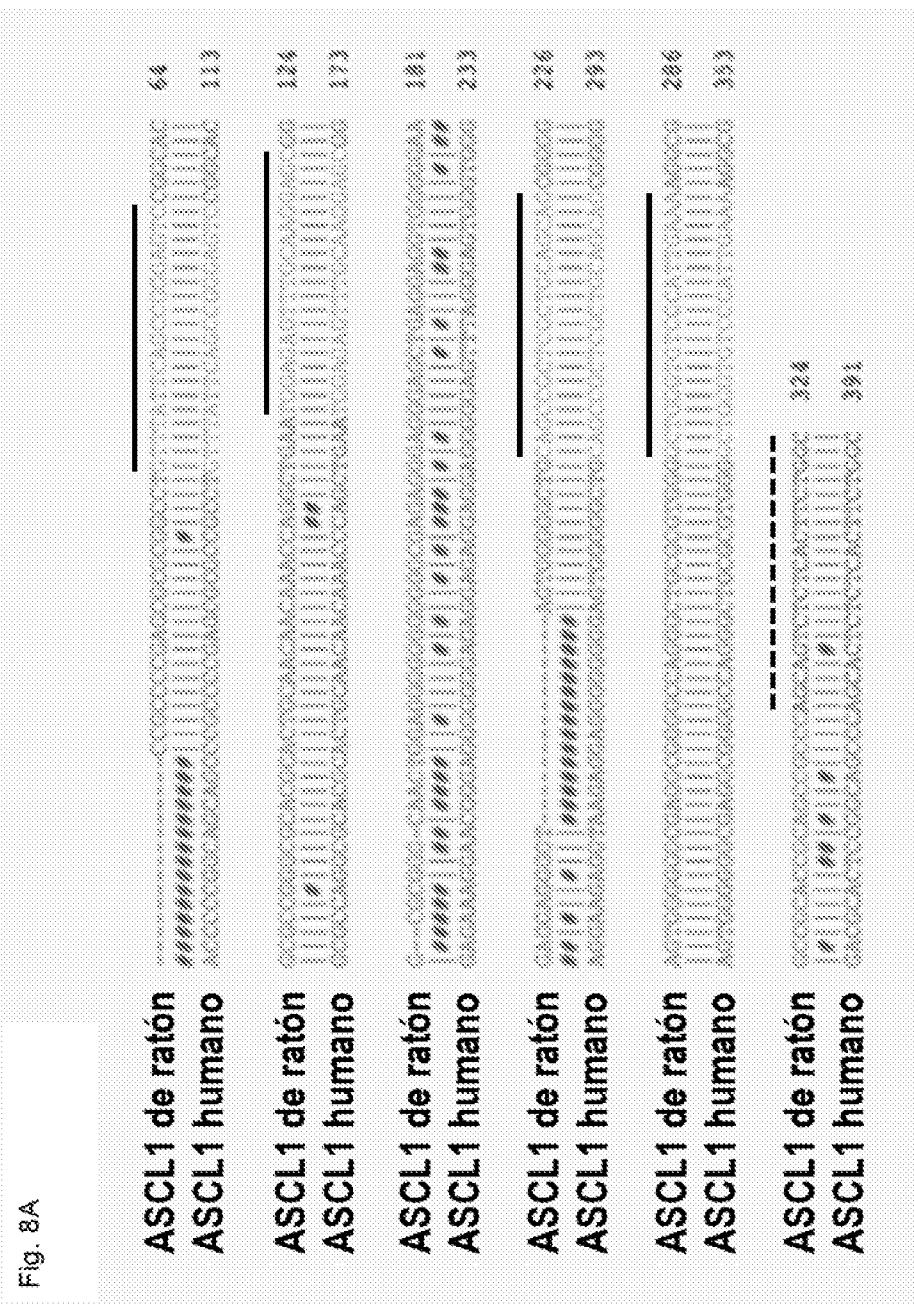


Fig. 8B

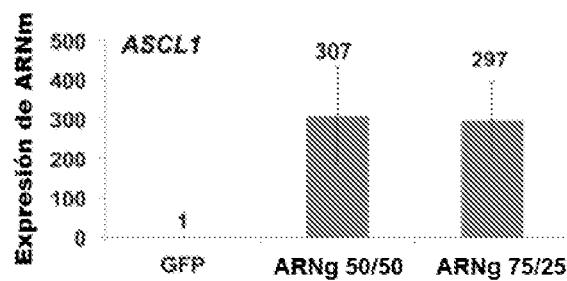


Fig. 8C

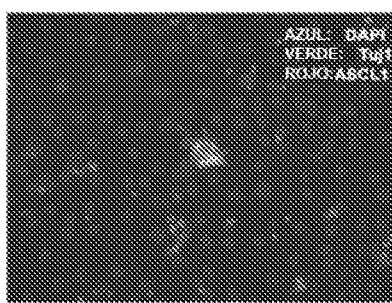


Fig. 8D



Fig. 8E

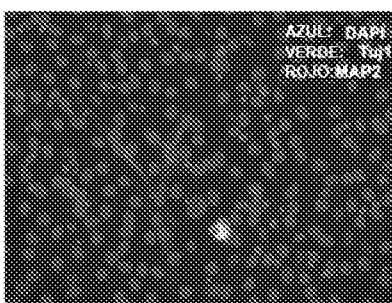


Fig. 8F

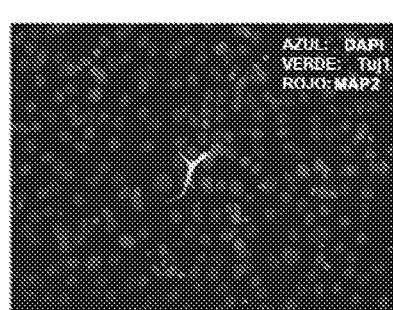
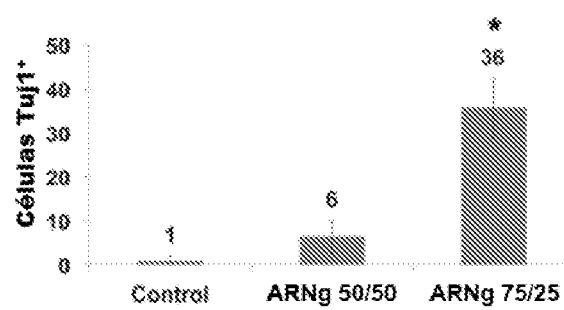


Fig. 8G



Fig. 8H



MDYKDHDGDYKDHDIDYKDDDKMAPKKKRKVGRGMDKKYSIGLAIGTNSVGWAVITD
EYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAETRLKRTARRRYTRRKNRICYLQ
EIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRILYALAHMIKFRGHFIEGDLNPNDSDVKLFQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNIALSLGLTPNFKSNFDAEDAKLQL
SKDTYDDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTITEIKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPPEKYKEIFFDQSNSKNGYAGYIDGGASQEEFYKFIPILEKMDGTEE
LLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRREKIEKILTFRIPY
YVGPLARGNSRFAMWTRKSEETITPWNEEVVDKGASAQSFIERNMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDOFLDNEENEDILEDIVLTTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRTGWGRRLSRKLINGIRDQSGKTIIDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQE
KLYLYLQNQGRDMYVDQELDINRLSDYDVDAIPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKNYWRQOLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSSRMNTKYDENDKLIREVKVTLKSCLVSDFRKDFQFYKVRREINNYHHAHDAY
LNAVVTALIKKYPKLESEFVYGDYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
!TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPVTAVSVLVAKVEKGKSKKLKSVKELLGITME
RSSFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRLMASAGELOKGNELALPSK
YVNFLYFLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL
AYNKHRDKPIREQAENIHLFTLTNLGAPAACKYFDTTIDRKRYTSTKEVLDATLHQSTITG
LYETRIDSQLGGDPIAGSKASKPKKRKVGRADALDDFDLDMGLSDALDDFDLDMGLGS
DALDDFDLDMGLSDALDDFDLMLINYPYDVPDYAS (SEQ ID NO: 1)

Etiqueta de epitopo FLAG = cursiva

Secuencia de localización nuclear = negrita

Cas9 de *Streptococcus pyogenes* = subrayado

VP64 (dominio VP16 mínimo 4x) = cursiva y negrita

Etiqueta de epitopo HA = cursiva y subrayado

Fig. 9A

GAGGGCCTATTCATGATTCCATATTCATATTCATACGATAACAAGGCTG
TTAGAGAGATAATTGAAATTAATTGACTGTAAACACAAAGATATTAGTACA
AAATACGTGACGTAGAAAGTAATAATTCTGGTAGTTGCAGTTTAAAA
TTATGTTTAAAATGGACTATCATATGCTTATCGTAACCTGAAAGTATTCGA
TTCTTGCTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTCGAGA
AGACCTGTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT (SEQ ID NO:2)

Promotor U6 = negrita

sitio de inicio de la transcripción +1 = subrayado

Sitios de restricción BbsI para clonar en ARN guía = cursiva y subrayado

Secuencia de ARN guía químérico = cursiva

Secuencia terminadora poli-T = negrita y subrayado

Fig. 9B

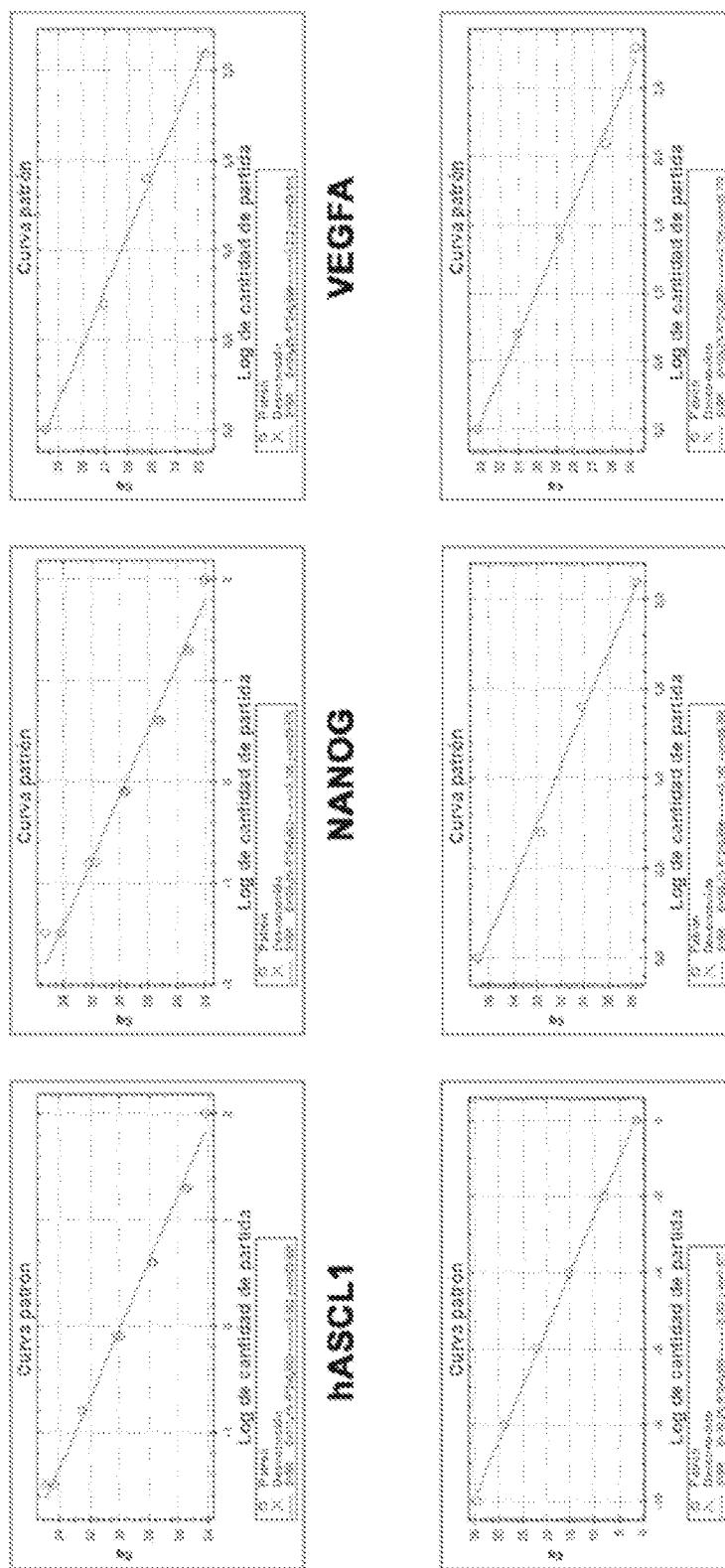


Fig. 10A

ES 2 987 399 T3

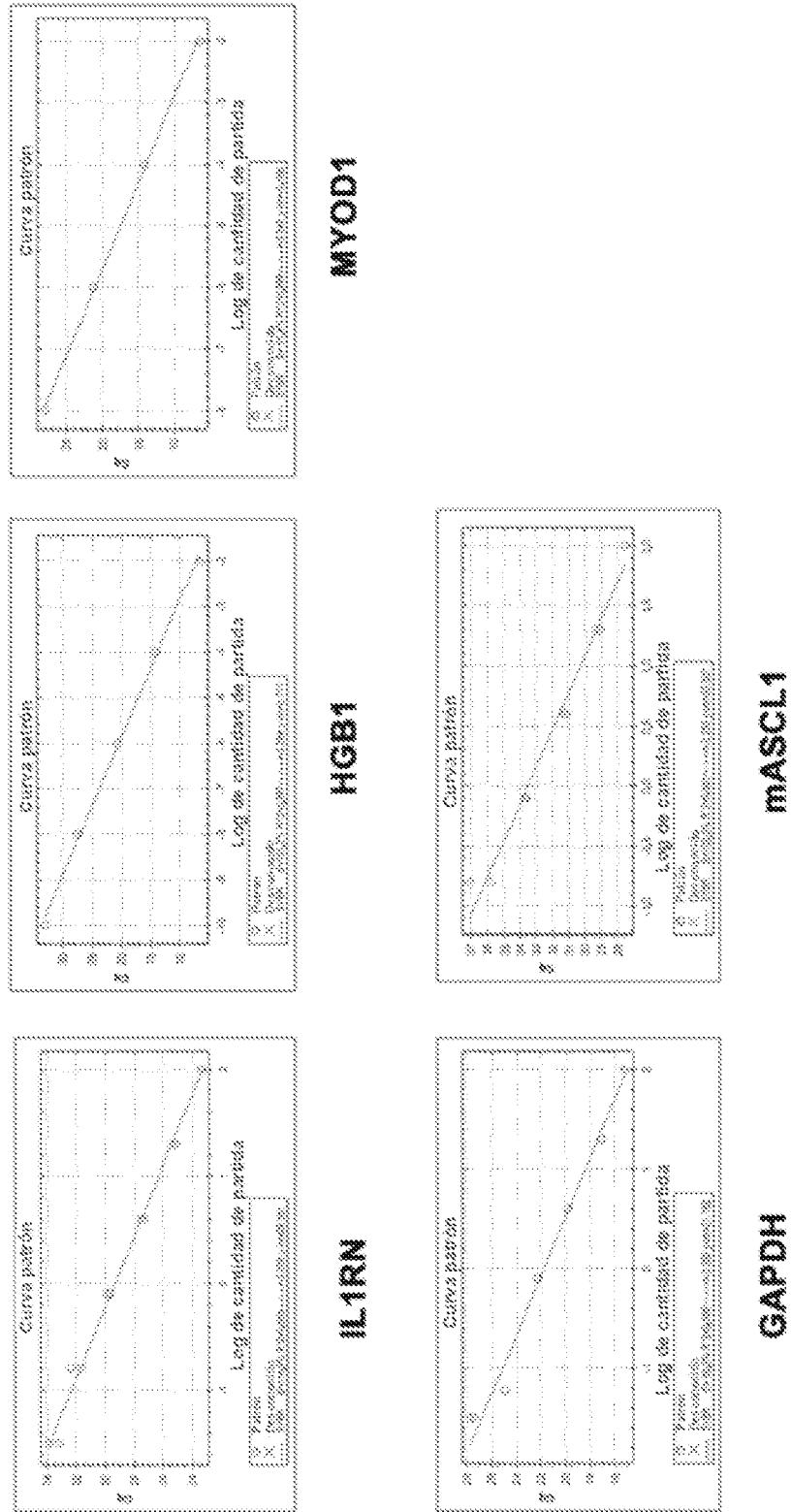


Fig. 10B

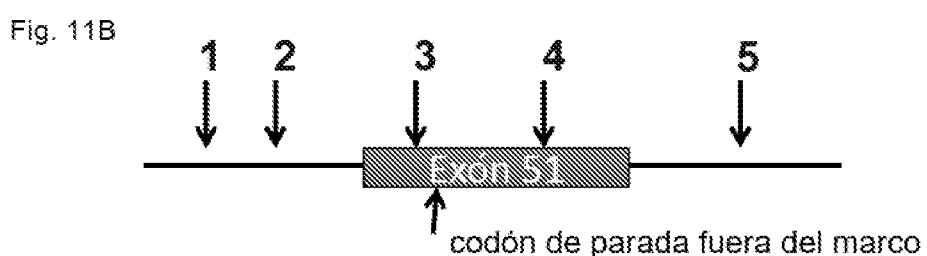
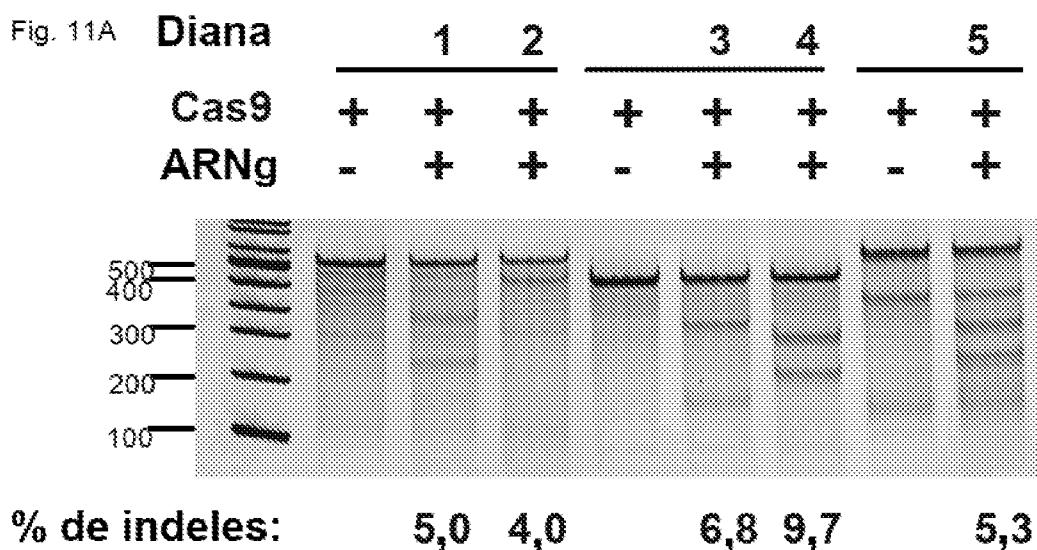


Fig. 11C

Tamaños de escisión esperados

CR1=352/237
 CR2=461/128
 CR3=301/150
 CR4=272/179
 CR5=280/203

ES 2 987 399 T3

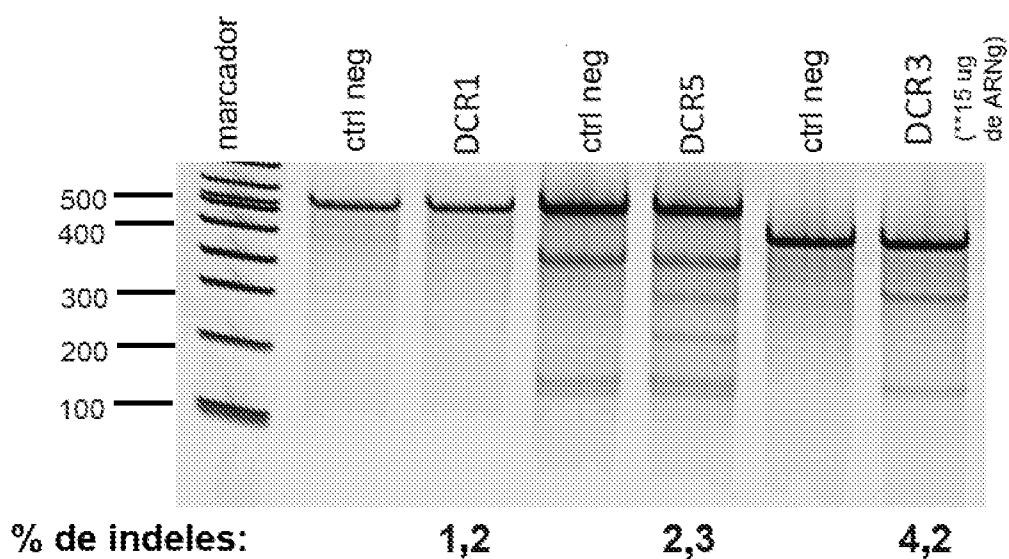


Fig. 12

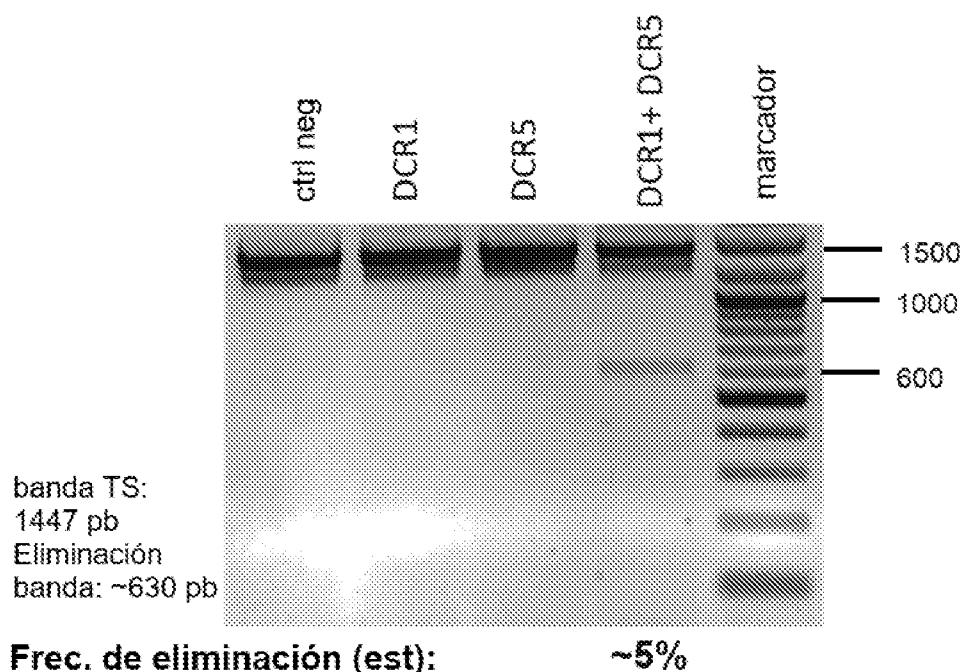


Fig. 13

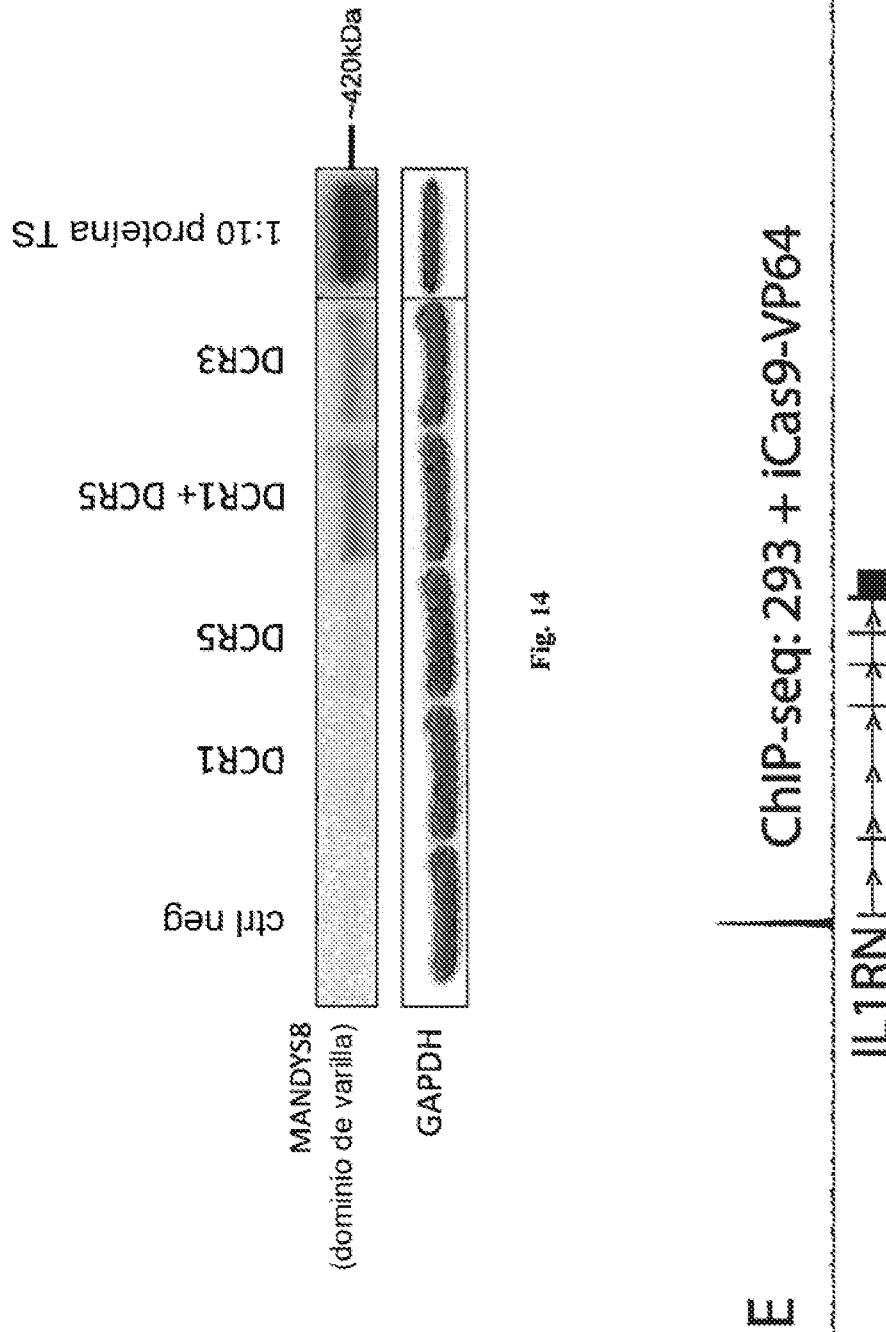


Fig. 16A

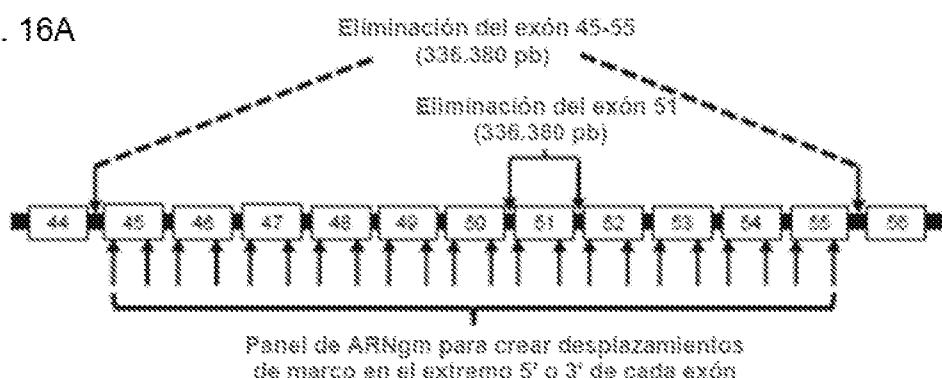


Fig. 16B

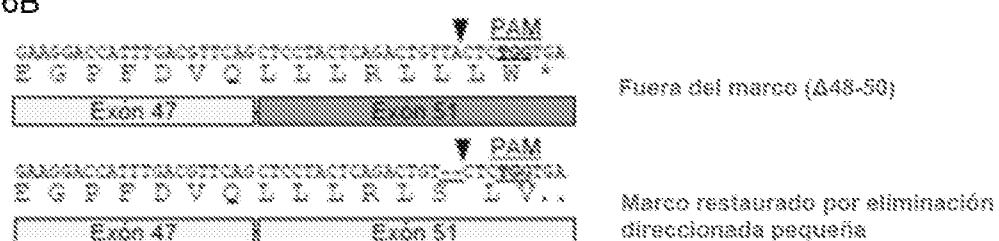


Fig. 16C

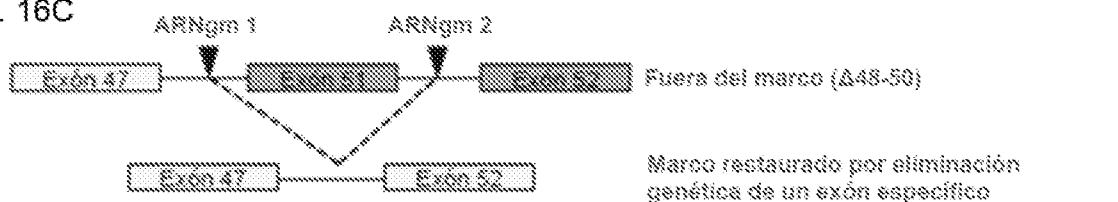
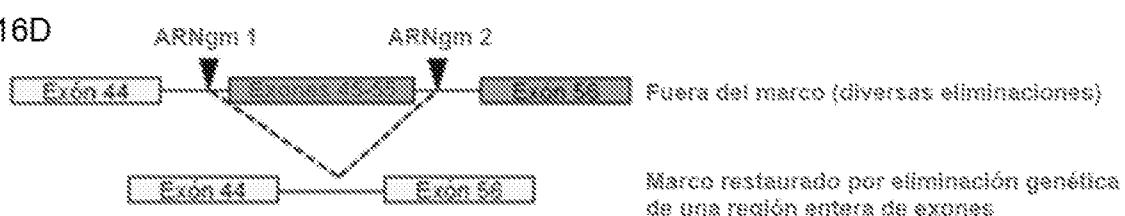


Fig. 16D



ES 2 987 399 T3

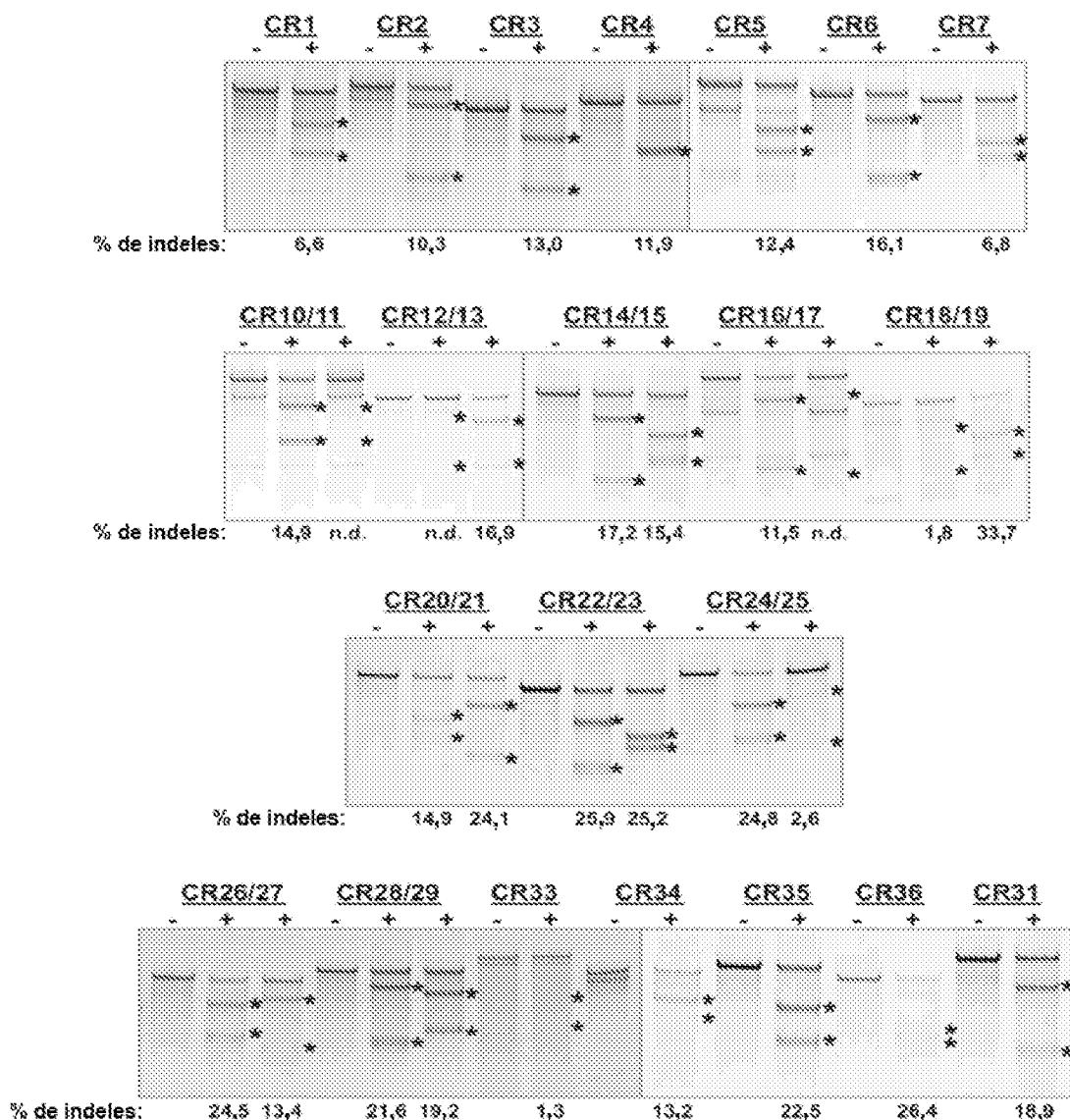


Fig. 17

ES 2 987 399 T3

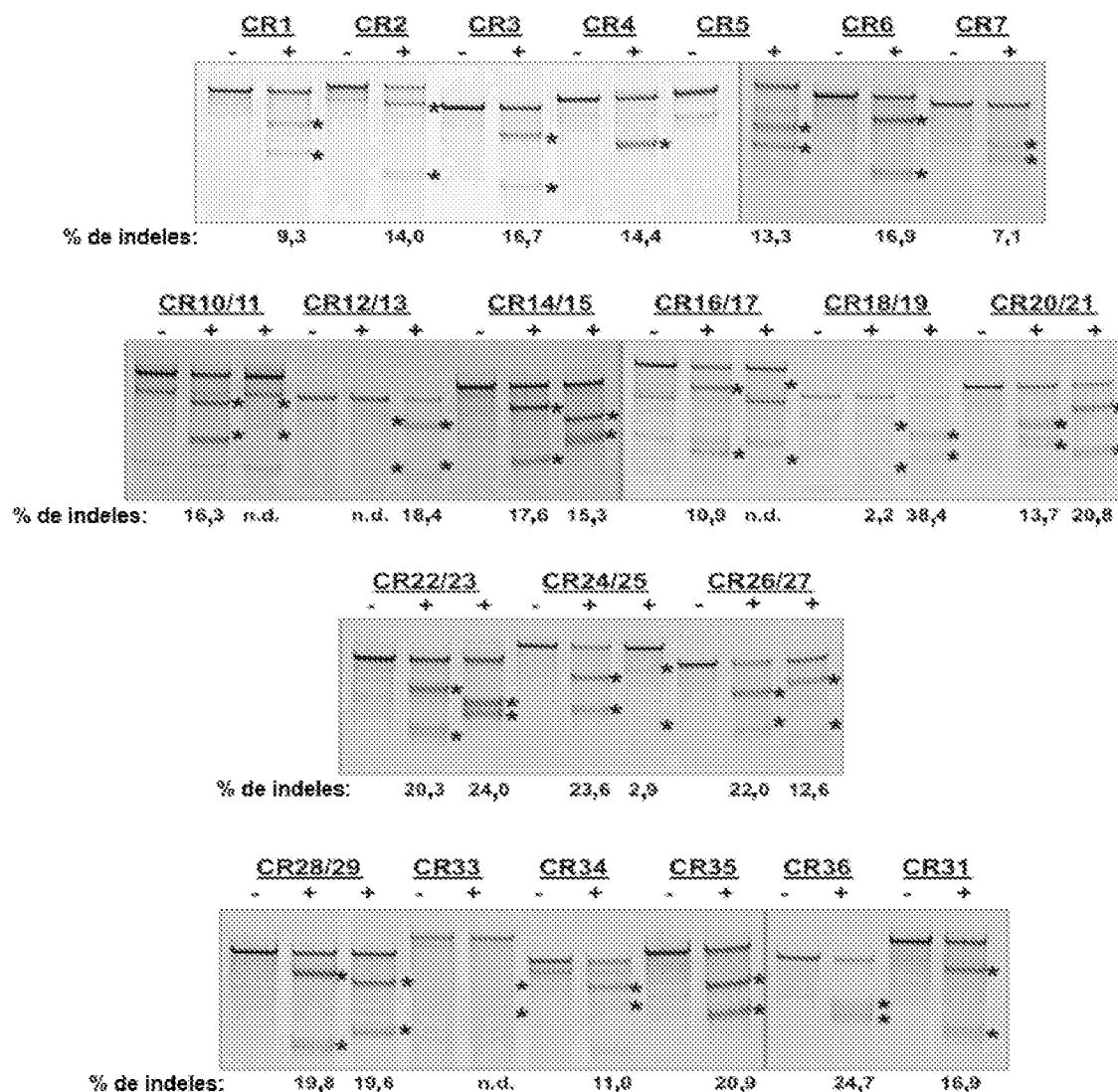


Fig. 18

Fig. 19A

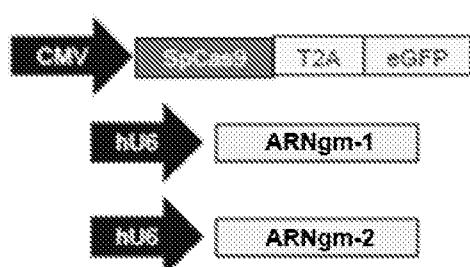


Fig. 19B

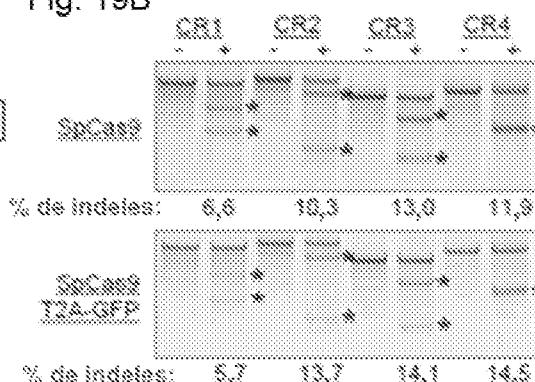


Fig. 19C

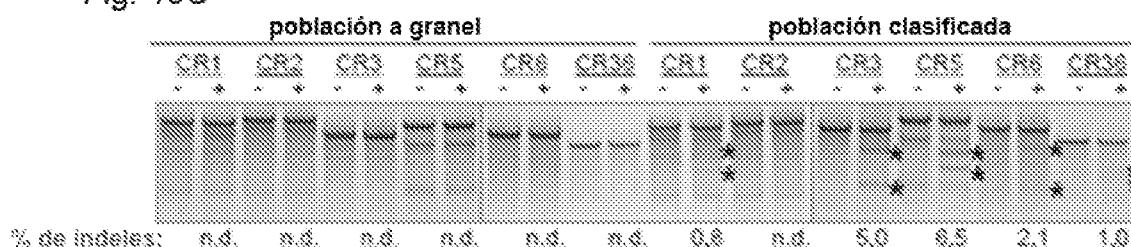


Fig. 19D

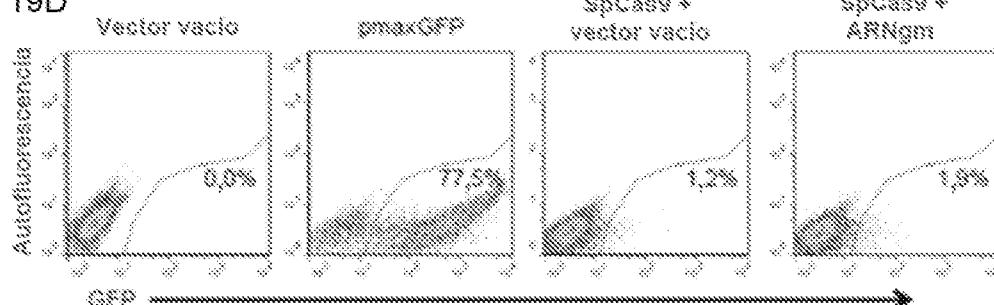


Fig. 20A

SEQ ID NO: Intón 50 Exón 51 PAM
 405 AAAATATTTAGCTCCCTACTCGACCTGTTACTCTTGTGACACAA
 406 TTTTATAAATCGACCAATTCTGACAAATGAGCCACTGTGTT
 ApnNm 5' ~ CCCUACUCAGACUCCAUACUC.....

Fig. 20B

Fig. 20C.

Eventos totales: 17/23 (52%)
Marco →1: 3/17 (18%)
Marco →2: 7/17 (41%)
Marco →3: 7/17 (41%)

Fig. 20D



Fig. 21A

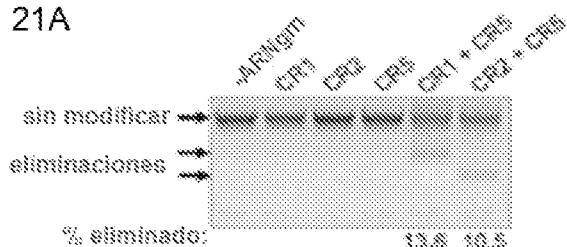


Fig. 21B



Fig. 21C

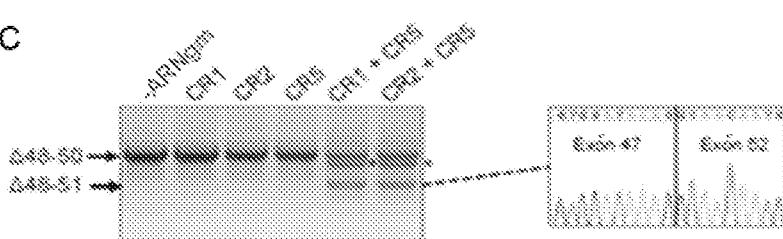


Fig. 21D



Fig. 22A

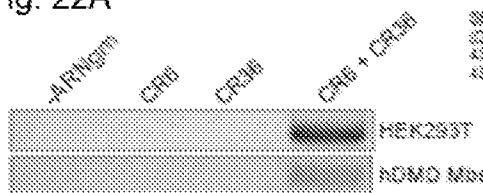


Fig. 22B

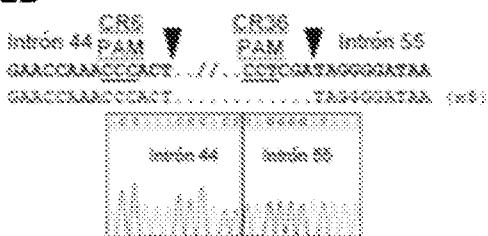


Fig. 22C

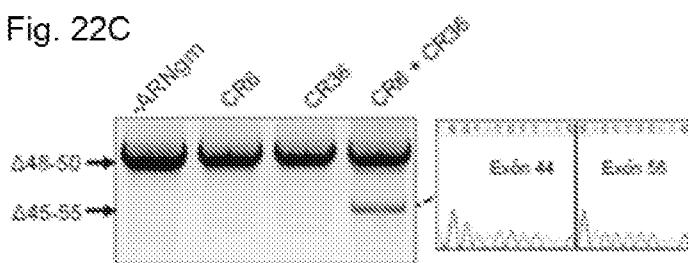


Fig. 22D

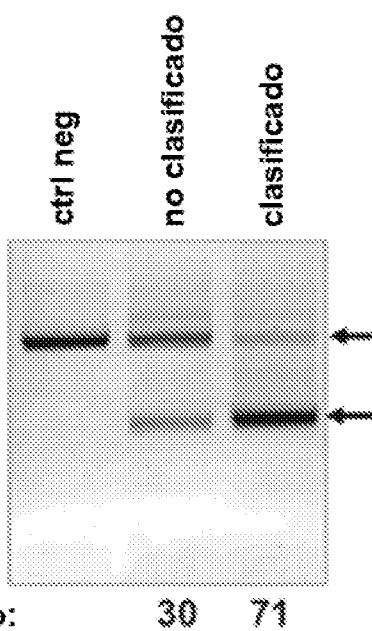


Fig. 23

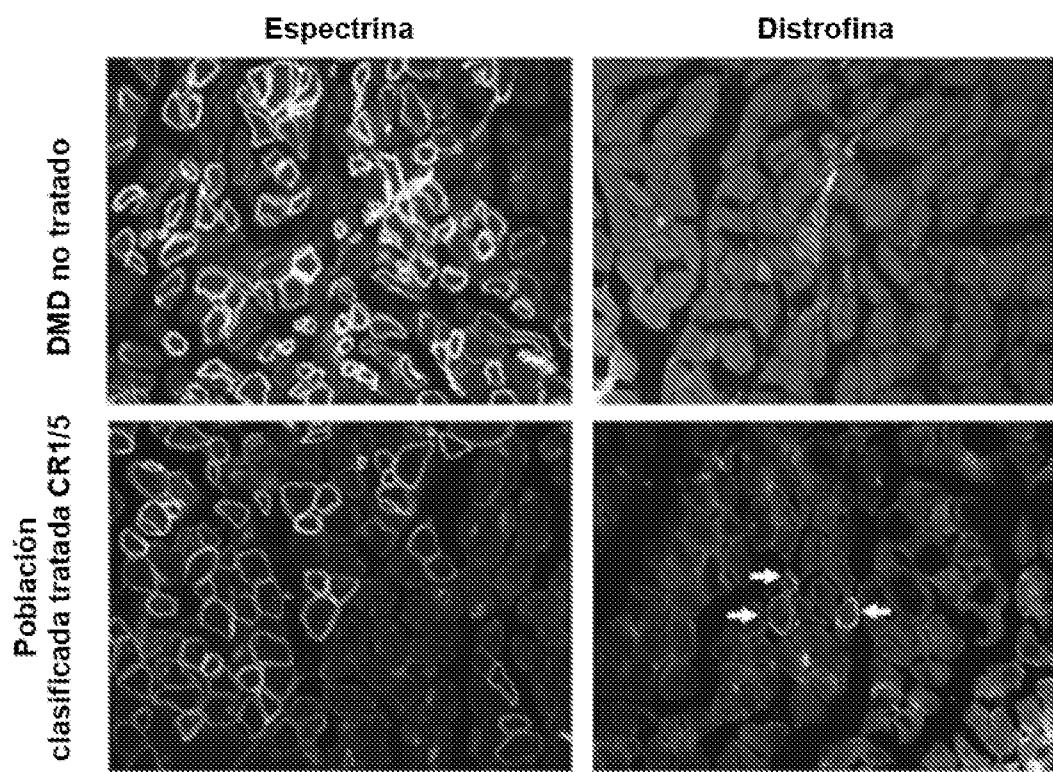


Fig. 24

ES 2 987 399 T3

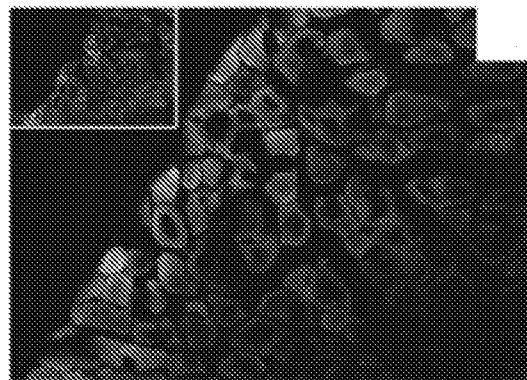


Fig. 25A

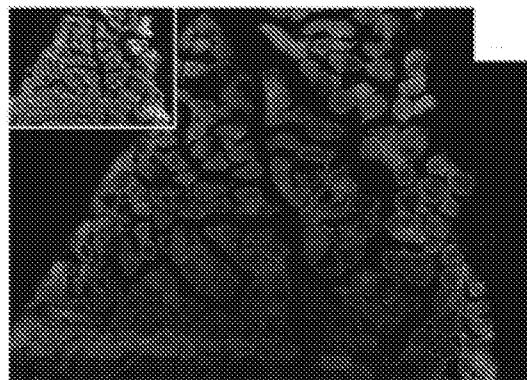


Fig. 25B

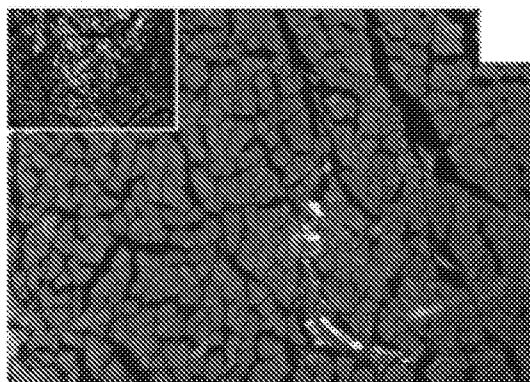


Fig. 25C

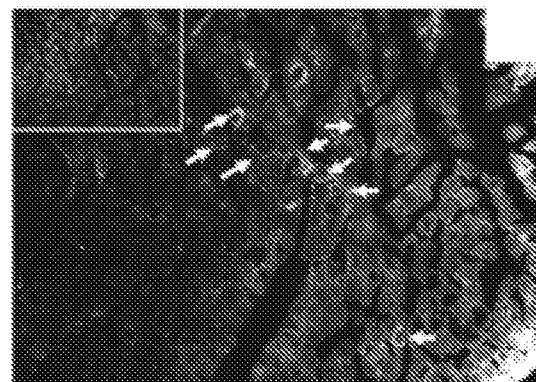


Fig. 25D

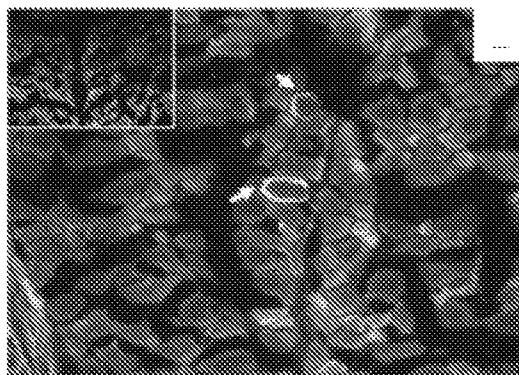


Fig. 25E

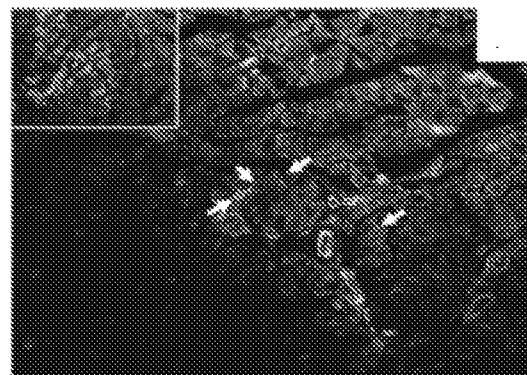


Fig. 25F

Fig. 26A

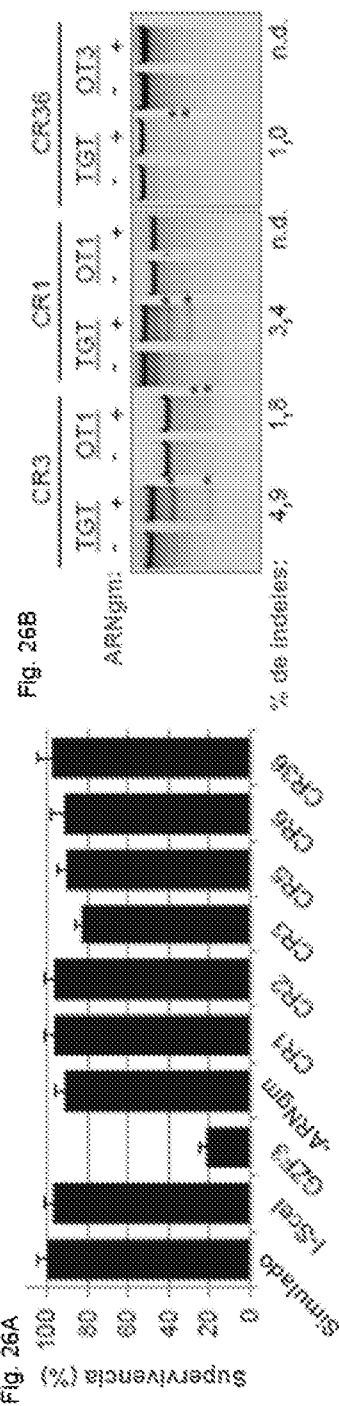


Fig. 26C

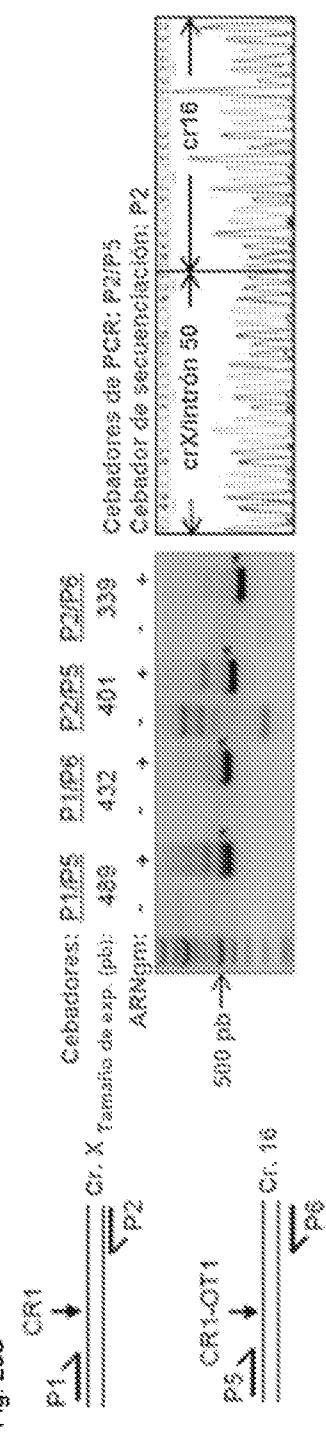
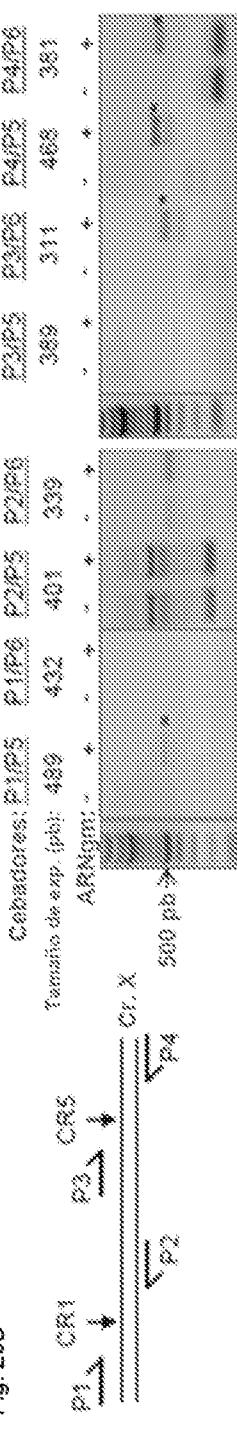


Fig. 26D



ES 2 987 399 T3

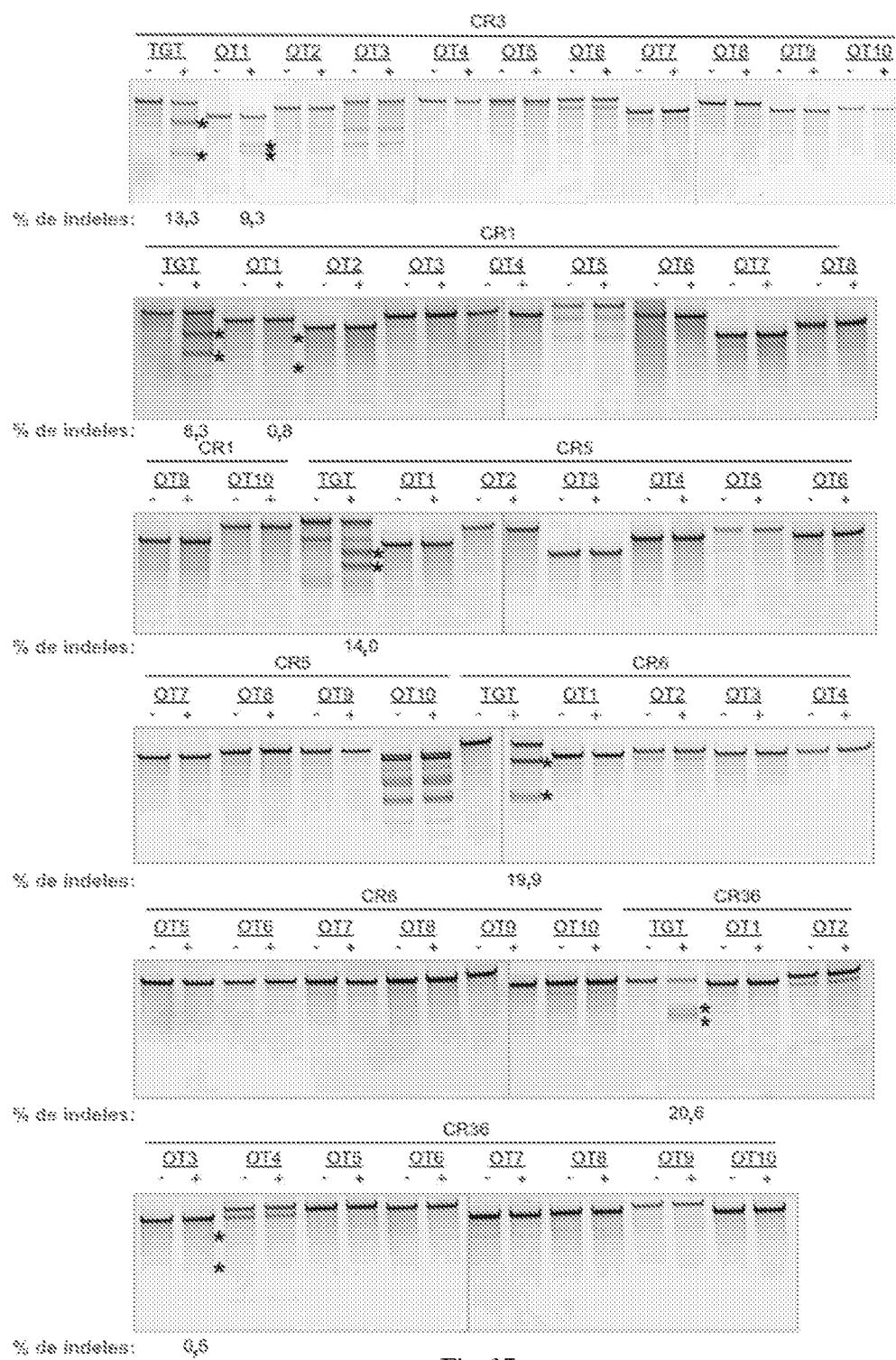


Fig. 27

Fig. 28A

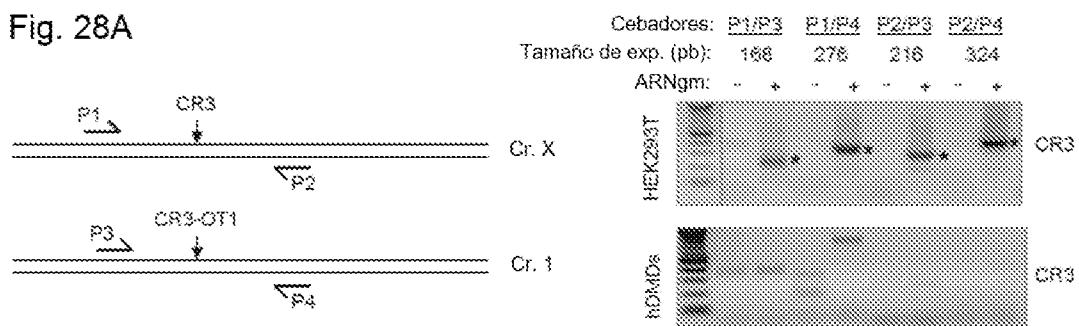


Fig. 28B

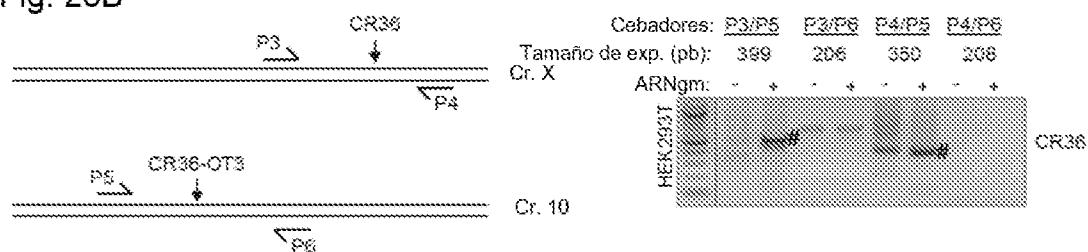
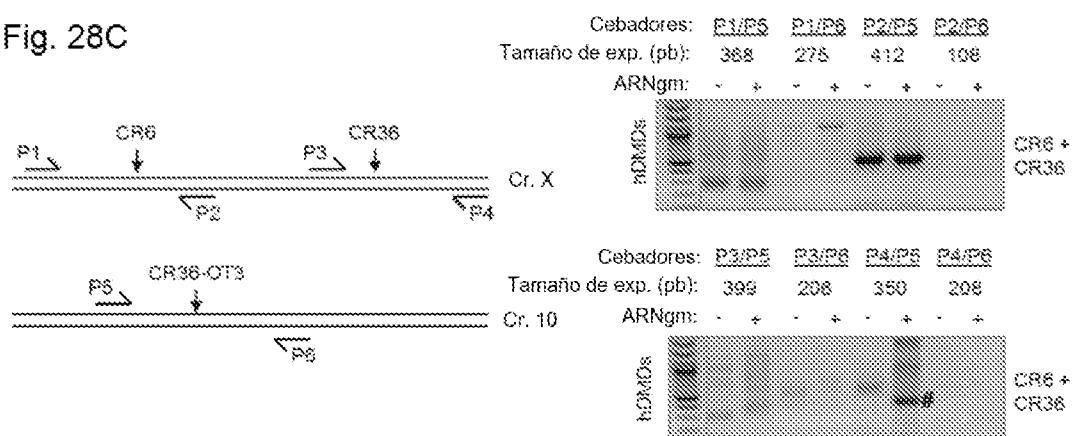


Fig. 28C



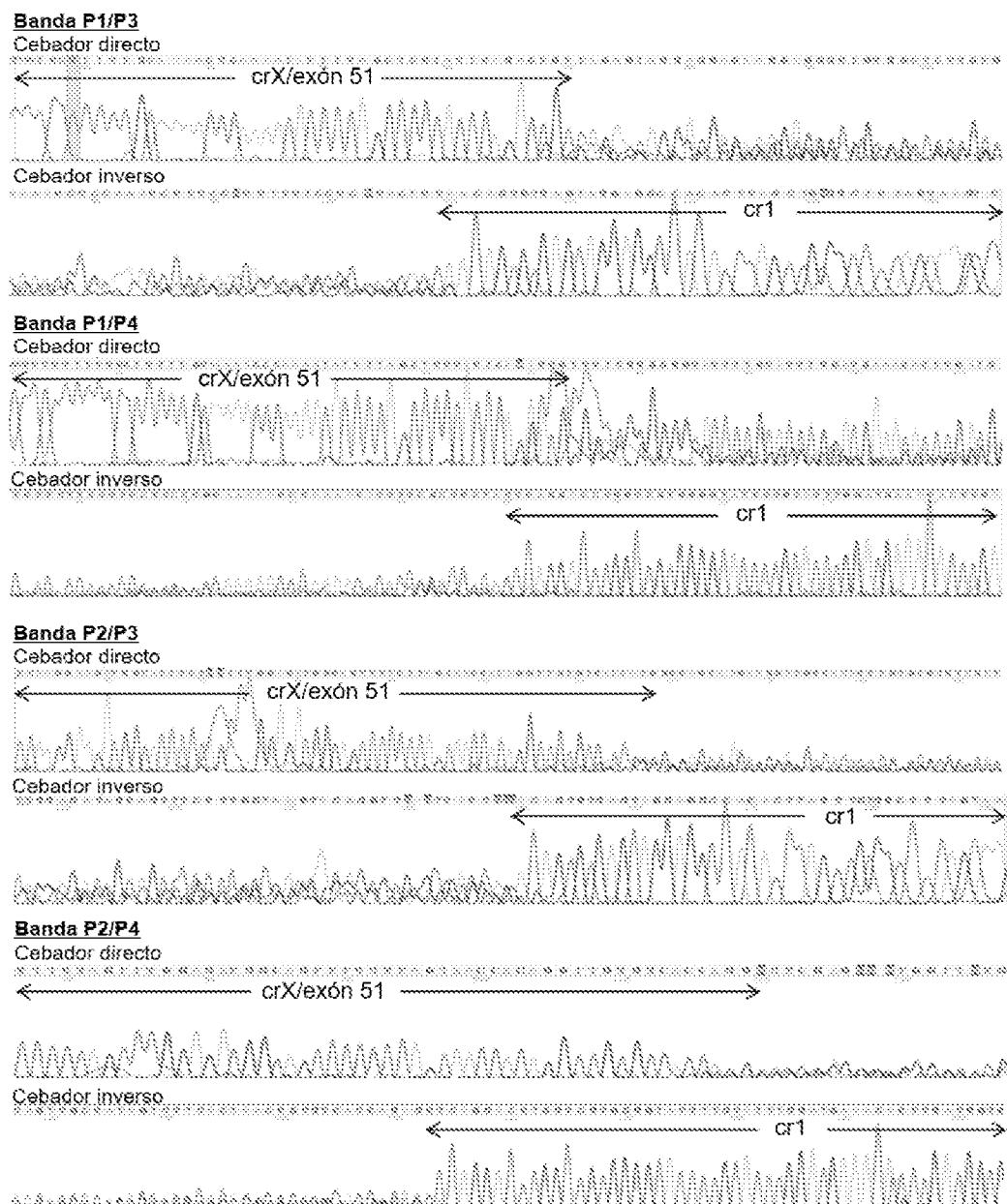


Fig. 29

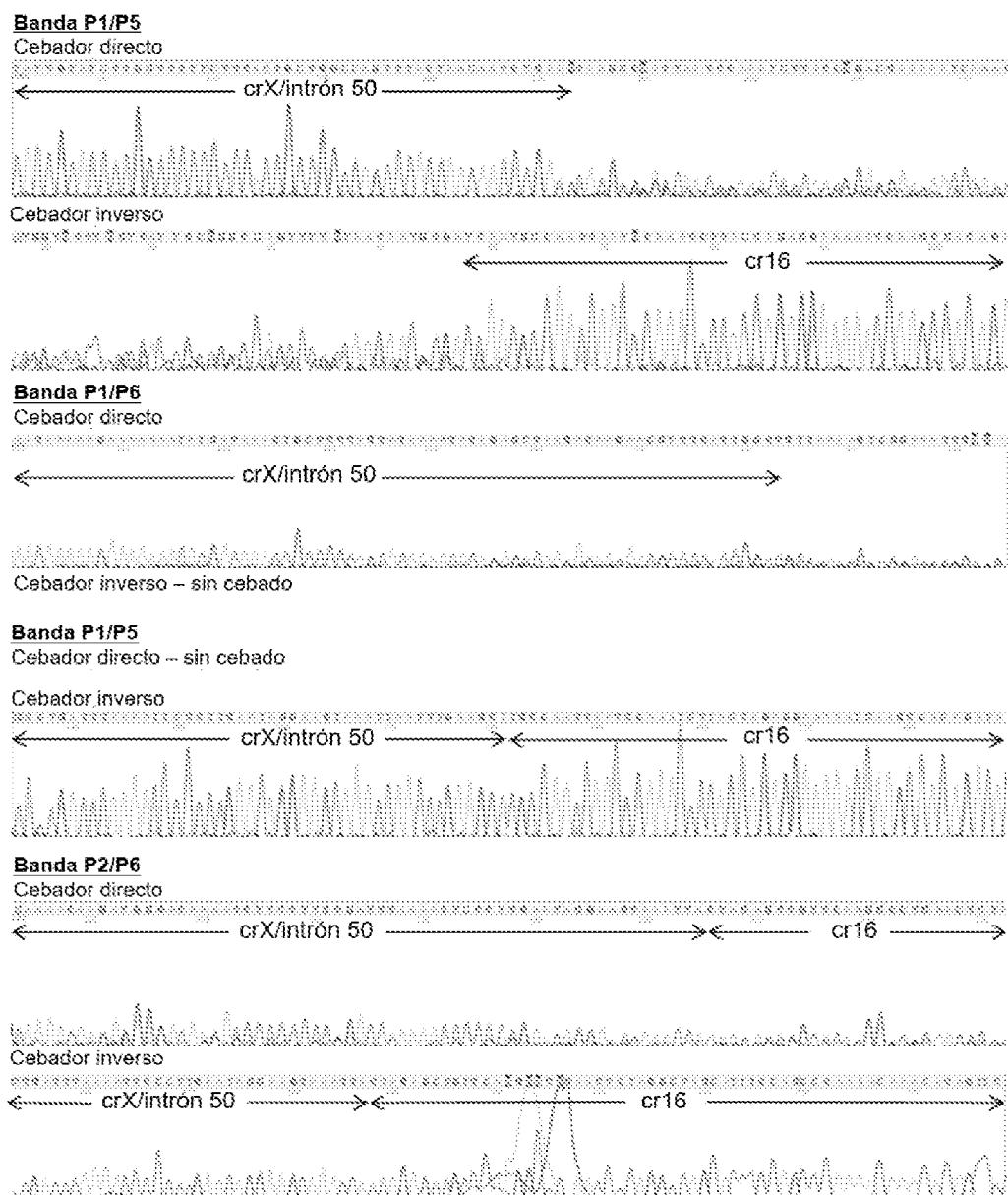


Fig. 30

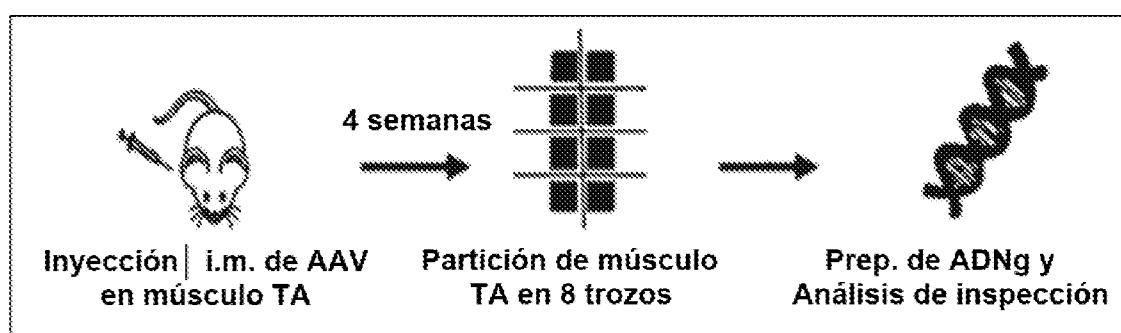


Fig. 31

Fig. 32A

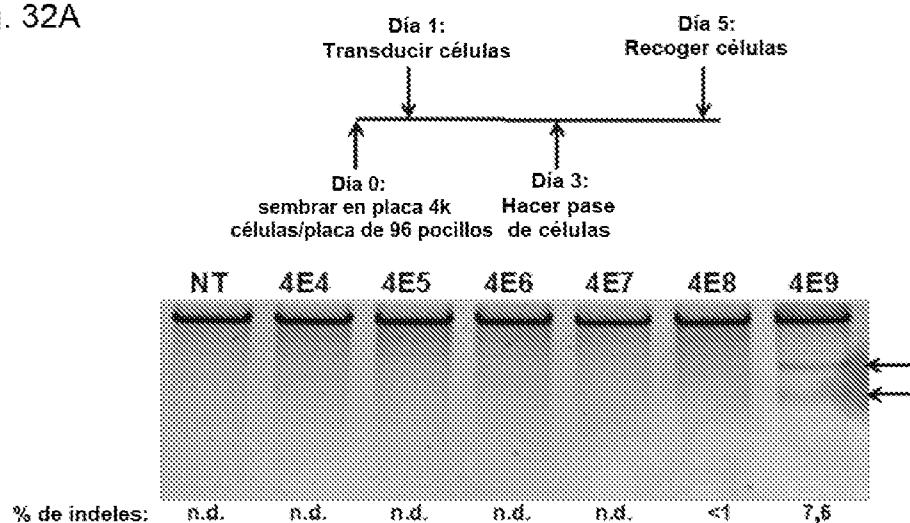


Fig. 32B

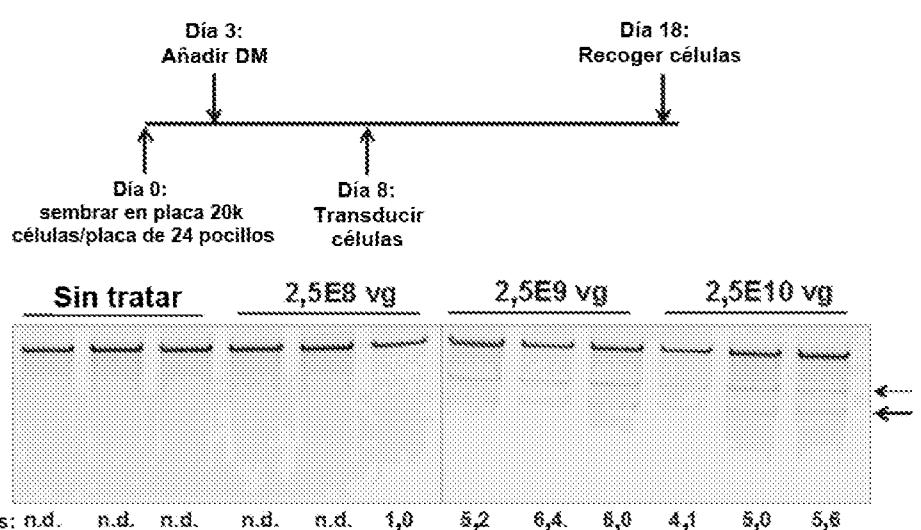
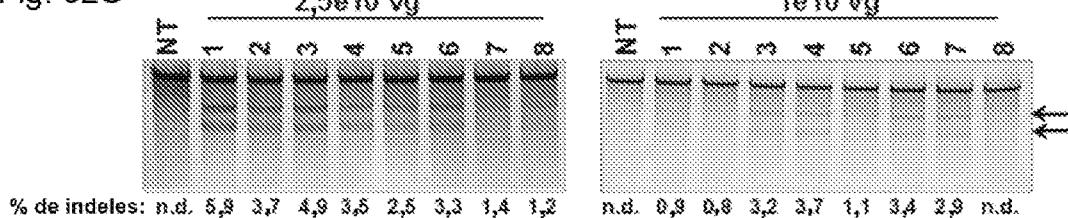


Fig. 32C



Secuencia de ADN Rosa T2A opt (SEQ ID NO: 434)

Secuencia de ADN Rosa T2A opt (SEO ID NO: 435)

MRSUYKDMUGDYKUNDIUXKNDIAAAMPYKHKRVGLQPEGGPKYKCPFGESFSUPGALVPHQKHTGEKF7KCPFGC
KSFPGQRALKLXHQPTATGCKPYKCPFGESFSUPGALVPHQKHTGEKF7KCPFGC
RAINKSELEKEKSEELANKLKYVPHETYLIELITLARNFTQDKILEMKVMRPPFMNYYGYCPGKHLGCGKKTIDGAIYTWTGSP
LDYGVUTKAYSGGYNLPIGQADEMENYVVEENYTRKNLNPNEWVYPSQSYTEFKFLFVSGHFKGNYKAQLTRIN
RTTNCNGAVLSEEVELLIGGENIKACTLTILEPVPKPNNGEINFEGCGSLLTCGDEEENPCPSSDVXINDGNYKDNDI
LYKDLUINMAPKKKKKVGLCPFGCKPYKCPFGESFSUPGALVPHQKHTGEKF7KCPFGC
GEKPYKCPFGESFSUPGALVPHQKHTGEKF7KCPFGC
NHQPTATGCKPYKCPFGESFSUPGALVPHQKHTGEKF7KCPFGC
HLNPNWVKVYPSSYTFRFLFVSGHFKGNYKAQLTRINHITNCNGAVLSEEVELLIGGENIKACTLTILEPVPKPNNGEINF

Fig. 33

Secuencia de cápside SASTG (SEQ ID NO: 430)

Fig. 34A

actggacaggtcagcgtggaaattgcgtggggactcacagaaaggaaacgcacaaacgttggaaatccagagattcagta
cacttccaaactacaacaagtctgttaatgtggactttactgttagacactaatgtgttatagtqaaacctccccca
ttggaaaccggatcttcacaaaggaaacttgtaaatctgttcaataaaccgtttaattcgttttcaggttgaacttt
ggctcttgtgcacttttatatatcttgtttccatgttactgcgttagataaagcagggcgtggggcgtggcgtggcgt
tcggcggttacacstyctgggtatatttaactctgcacatacttcgttgatgtgggttggccactccctatccgg
cactcgctcgctcgatggggccgcacutgcacgtcgttgcgcaccccccgttggtggccggcccccacccggcgg
ggcggagggcgtataggggggtggccaa

Secuencia peptídica de cápside SASTG (SEQ ID NO: 437)

MSAIGYLPIQLEDDILSEGIPEWALKRFGVIEOPKAKQQDNNRGIVLPGYNYLGPQNGILDKEDPVNEADAALEHCK
AYDQQLMAGINIPYLKYNHADAEFQEKLOKLTETFOGNLGRAVUQANKRILLEPLGIVINERAAKTAPOKNERPVDCQFQKFD
SSSGVYKDNQPARKEIINFOQFTLGESKVPIPQFMEPPFAAPTSLGSTNTMASGGAGMAIDNTEGIVNLSGGONWHCD
SQWLGDNVITTCRTTWALPTYNNHLYQYSSASTGAGNDNHYFGYCTFWGYFLPNRPHCNPSPRNQPLINNNWDFR
PERLSEKLFNIQVEVTQNDVTTIANNLITSTQVFTDSEYQLPYVILGSANQCOLPPFFAVTPNVPQYGYTLLENGS
QAVGRSSPFYCLSYFPQQLMKTGANNQFSYTFKUVPRHCSYAHQSLRILMNPLIDQYLYTLMRTQGTTSQITTNQCKL
LPSQAGHQMSLQARWLLPGFCYKQORLQKNTANDNNNNPWTAAACKYHUNGRDLVNPQPARASHKDEEERSTPMSR
GNLIPGKGEGITAGSNALLDNVMTLRRRTRITNPVATEQYGTVANNLQGENTAPTTNTVNDQGALPGCNVWQDRDVYLO
GPIWAKLPHIDQHPTSFIMUGPULRHPPTQIMIKNTFPVPAHPPFTFSPAKFASFIHQYDTCGVSVELIEXELQENS
KPNWPSIQYISNYNKSVNYDFTVLTNGVYSEPRPITGTRYLTFL

Fig. 34B

Sitio diana DZF16 ZFN (SEQ ID NO: 442):

5' -CAAACTAGAAATGCCAICCTCCTTGAACTGGGAACTACACTCCATGGACG-3'
 3' -GTTTGAATCTTACGGTAGAAGGAACTACACACCTCCATGGACG-5'

Secuencia de aminoácidos completa izquierda DZF16-L6 (SEQ ID NO: 443)

M DY K N H G D Y K D N D I D Y K D D D K M A P K K R K V G R L E P G E K P Y K C P E C G K S F S Q R K D A L P G H Q R T H T G E K P Y
 K C P E C G K S F S H E T T L T N H Q R T H T G E K P Y K C P E C G K S F S Q R N A L A G H Q R T H T G E K P Y K C P E C G K S F S H K N A
 L Q N H Q T H T G E K P Y K C P E C G K S F S U P G H I V R H Q R T H T G E K P Y K C P E C G K S F S T S Q N L V D H Q R T H T G A A A R
 A L V K S E L E E K K S E L R R K L K Y V P R E Y T E L I E I A R N P T Q D R I L E M K V M E F F M K V Y G Y R G E H L P G S P K P D G A I
 Y T V G S P I D Y G V I V D T K A Y S G G Y N L P I G Q A D E M E R Y V E E N O T R D K H L N P N E W W K V Y P G C V T E F K P L F V S G H
 F G G R Y K A Q L T R L N R I T N C N G A V I L S V E E L L I G G E M I K A G T I T L E E V R R K F N N G E I N F *

Secuencia de aminoácidos completa derecha DZF16-R6 (SEQ ID NO: 444)

M DY K N H G D Y K D N D I D Y K D D D K M A P K K R K V G R L E P G E K P Y K C P E C G K S F S Q R S L V G H Q R T H T G E K P Y
 K C P E C G K S F S H E T T L R B Q R T H T G E K P Y K C P E C G K S F S T S G H I V R H Q R T H T G E K P Y K C P E C G K S F S Q R A R
 A L V K S E L E E K K S E L R R K L K Y V P R E Y T E L I E I A R N P T Q D R I L E M K V M E F F M K V Y G Y R G E H L P G S P K P D G A I
 Y T V G S P I D Y G V I V D T K A Y S G G Y N L P I G Q A D E M E R Y V E E N O T R N K H I N P N E W W K V Y P G C V T E F K Y L F V S G H
 F G G N Y K A Q L T R L N R I T N C N G A V I L S V E E L L I G G E M I K A G T I T L E E V R R K F N N G E I N F *

Fig. 35

Sitio diana E51C3 (SEQ ID NO: 445):

5' ~ (c) ATGTGCGCATGACTGGCGCGGG (a) ~ 3'
 3' ~ (a) TAGACGGGTACTGACCGCGTCGC (t) ~ 5'

Secuencia de aminoácidos completa izquierda E51C3-3L (SEQ ID NO: 446)

M DY K N H G D Y K D N D I D Y K D D D K M A P K K R K V G R L E P G E R P F Q C R I C M R N F S K Q A L A V H T P T H T G E K P F
 Q C R I C M P N F S Q S T T L K R H I R T H T G E K P F Q C R I C M R N F S R S D R L S L H I K T H L G S Q U V S E L E E K K S E L R H
 K L K Y V P R E Y T E L I E I A R N P T Q D R I L E M K V M E F F M K V Y G Y R G E H L P G S P K P D G A I Y T V G S P I D Y G V I V D T K
 A Y S G G Y N L P I G Q A D E M E R Y V E E N O T R D K H L N P N E W W K V Y P G C V T E F K Y L F V S G H F G G N Y K A Q L T R L N R S T
 N C N G A V I L S V E E L L I G G E M I K A G T I T L E E V R R K F N N G E I N F *

Secuencia de aminoácidos completa derecha E51C3-3R (SEQ ID NO: 447)

M DY K N H G D Y K D N D I D Y K D D D K M A P K K R K V G R L E P G E R P F Q C R I C M R N F S Q R A L O N H T R T H T G E K P F
 Q C R I C M R N F S Q S T T L K R H I R T H T G E K P F Q C P I C M R R N F S D G H L T R L G S Q U V S E L E E K K S E L R H
 K L K Y V P R E Y T E L I E I A R N P T Q D R I L E M K V M E F F M K V Y G Y R G E H L P G S P K P D G A I Y T V G S P I D Y G V I V D T K
 A Y S G G Y N L P I G Q A D E M E R Y V E E N O T R D K H L N P N E W W K V Y P G C V T E F K Y L F V S G H F G G N Y K A Q L T R L N R S T
 N C N G A V I L S V E E L L I G G E M I K A G T I T L E E V R R K F N N G E I N F *

Fig. 36

Sitio diana DZF15 (SEQ ID NO: 448):

5' -ACTAGAAAATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCT-3'
 3' -TGATCTTTACGGTAGAAGGAACATCACACCTCCATGGACGAGA-5'

Secuencia de aminoácidos completa izquierda DZF15-L6 (SEQ ID NO: 449)

MDYKIDHEDGYKDHDIDYKUDUDDKMAPKKKKRKVGRLEPGEKPYKCPECGKSFSDNPTTLNHNQNTHTGEKPY
 KCPECGKSFSDNQNALAGHQPTHTGEKPYKCPECGKSFSDNQNALQNEQHTHGKPTHTGEKPYKCPECGKSFSDPGL
 LVRHQNTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDNQNLVNRHQNTHTGAAAR
 ALVKSELEEKKSSELPHKLKYVPHHEYIELTIEIARNPTQDRILEMKVMEFTMNVYGYRGEHLGGCRKPTDGAI
 YTVGSPIDYGVIVDVKAYSGGYNLPIQQADEMERYVTKENQTRNKHISPNENWKVYPSCVTEFYTFLFVSGH
 FKGNYKAQJLTPLNRTKNONGAVLSVEELLLIGGEMIKAGTLTLEEVRPKTNNGEINF*

Secuencia de aminoácidos completa derecha DZF15-R6 (SEQ ID NO: 450)

MDYKIDHEDGYKDHDIDYKUDUDDKMAPKKKKRKVGRLEPGEKPYKCPECGKSFSDNQNALAGHQPTHTGEKPY
 KCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDPGL
 LVRHQNTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGAAAR
 ALVKSELEEKKSSELPHKLKYVPHHEYIELTICIAARNPTQDRILEMKVMEFTMNVYGYRGEHLGGCRKPTDGAI
 YTVGSPIDYGVIVDVKAYSGGYNLPIQQADEMERYVTKENQTRNKHISPNENWKVYPSCVTEFYTFLFVSGH
 FKGNYKAQJLTPLNRTKNONGAVLSVEELLLIGGEMIKAGTLTLEEVRPKTNNGEINF*

Secuencia de aminoácidos completa izquierda DZF15-L5 (SEQ ID NO: 451)

MDYKIDHEDGYKDHDIDYKUDUDDKMAPKKKKRKVGRLEPGEKPYKCPECGKSFSDNQNALAGHQPTHTGEKPY
 KCPECGKSFSDNQNLQNEQHTHGKPTHTGEKPYKCPECGKSFSDNQNLQNEQHTHGKPTHTGEKPYKCPECGKSFSDNQNL
 LVRHQNTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGAAARALVKSELEEKKSSELPHKLKYVPHHEYIEL
 IEIARNPTQDRILEMKVMEFTMNVYGYRGEHLGGCRKPTDGAIYTVGSPIDYGVIVDVKAYSGGYNLPIQQA
 DEMERYVTKENQTRNKHISPNENWKVYPSCVTEFYTFLFVSGHFKNYKAQJLTPLNRTKNONGAVLSVEEL
 LLIGGEMIKAGTLTLEEVRPKTNNGEINF*

Secuencia de aminoácidos completa derecha DZF15-R5 (SEQ ID NO: 452)

MDYKIDHEDGYKDHDIDYKUDUDDKMAPKKKKRKVGRLEPGEKPYKCPECGKSFSDNQSLVNRHQPTHTGEKPY
 KCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGEKPYKCPECGKSFSDNQNL
 LEPHQNTHTGEKPYKCPECGKSFSDNQNLVNRHQPTHTGAAARALVKSELEEKKSSELPHKLKYVPHHEYIEL
 IEIARNPTQDRILEMKVMEFTMNVYGYRGEHLGGCRKPTDGAIYTVGSPIDYGVIVDVKAYSGGYNLPIQQA
 DEMERYVTKENQTRNKHISPNENWKVYPSCVTEFYTFLFVSGHFKNYKAQJLTPLNRTKNONGAVLSVEEL
 LLIGGEMIKAGTLTLEEVRPKTNNGEINF*

Fig. 37

Sitio diana E51C-4 (SEQ ID NO: 453):

5' ~ (t) GCCATCTTCCTTGATGTTGGAGGT (a) ~3'

3' ~ (a) CGGTACGAAAGRACTACGACCTCCA (t) ~3'

Secuencia de aminoácidos completa izquierda E51C-4L (SEQ ID NO: 454)

MDYKNDGUYKNCIDIDYKDDDKNAFKRKREVVRLEPGCRFPFOCRICMRNFSSESKLARHTRHTGCNPFFQCRICMRNFSVYRNLTRHLATHTGEKPFQCRICMRNFSQENLGRNLXTRTGAARAALVKGELLEEKKSELRRKILKVVPHHEYIELIIZIARNPTQDRILEMKVMESTPNKVYCYRGERHLOGSRKPOGAIYTVCSPIDYGVIVDTKAYSGGYNLPICQACDEMORYVKENQTRNKHINPNEWWKVYPSSVTTEKFPLFVSCHIFXONTKAQLTRLNPIITNCNGAVLSVEELLIGSEMIIKGTLTLEEVREKKENNGEINF*

Secuencia de aminoácidos completa derecha E51C-4R (SEQ ID NO: 455)

MDYKNDGUYKNCIDIDYKDDDKNAFKRKREVVRLEPGCRFPFOCRICMRNFSIPNIKALARHTRHTGCNPFFQCRICMRNFSQENLGRNLXTRHKTHTGAARAALVKGELLEEKKSELRRKILKVVPHHEYIELIIZIARNPTQDRILEMKVMESTPNKVYCYRGERHLOGSRKPOGAIYTVCSPIDYGVIVDTKAYSGGYNLPICQACDEMORYVKENQTRNKHINPNEWWKVYPSSVTTEKFPLFVSCHIFXONTKAQLTRLNPIITNCNGAVLSVEELLIGSEMIIKGTLTLEEVREKKENNGEINF*

Fig. 38

"Vector individual, sistema CRISPR múltiple"



"Vector doble, sistema CRISPR múltiple"



"Vector individual, sistema ARNg individual"

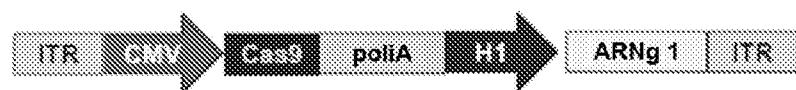


Fig. 39

SaCas9-NLS (NLS subravado) (SEQ ID NO: 459)

ARNg SaCas9 (SEQ ID NO: 460)

Fig. 40

NmCas9 (NLS 1 subrayado NLS 2 subrayado/negrita, etiqueta HA negrita) (SEQ ID NO: 461)

Horquilla corta NmCs9 de Thomson PNAS 2013 (SEQ ID NO: 462)

[peachleafbinder] GTTGTGGCTCCCTTCCTCATTTGAAACCGATGAGACCGCTTGCTACAAATAGCGCGCTGTA
KAGATGTCGCGCAGCTCTGCCTTAAACGCTTCTGTTTAAAGCTTCTGTTT

Horquilla larga NmCs9 de Church Nature Biotech 2013 (SEQ ID NO: 463)

[untranslated] TTTTGTAATCTGCCCTTCATTCGCGCTGCTACAAAGAAAGTTGTCGCAACTGCCTATGAGAACCGTTGCTACAAATAAGGCCGCTGTGAAAAGATGTCGCCGCAAGGCTCTGCCCTTAAGGCTCTGCTTAAAGGGGTTTTTCTT

Fig. 41

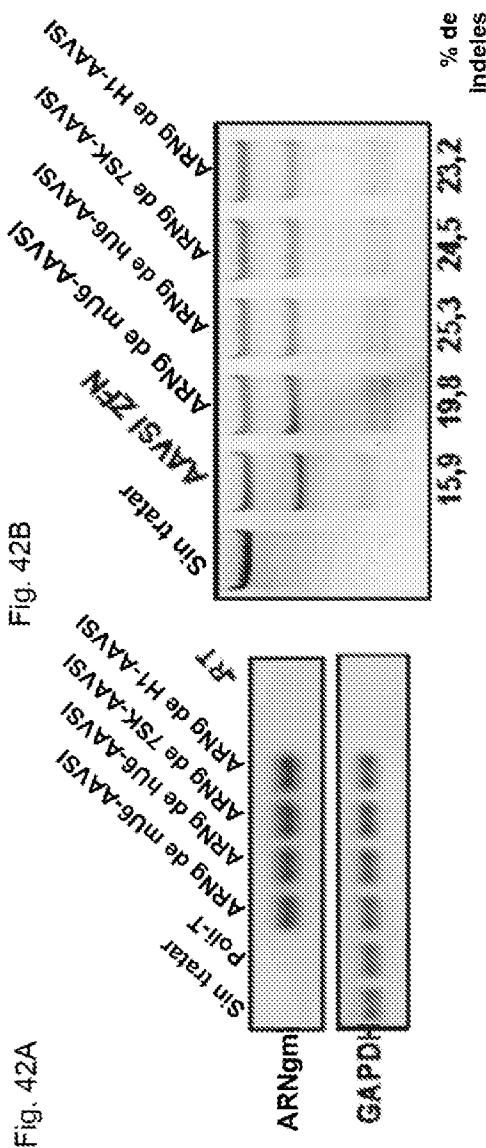


Fig. 42A

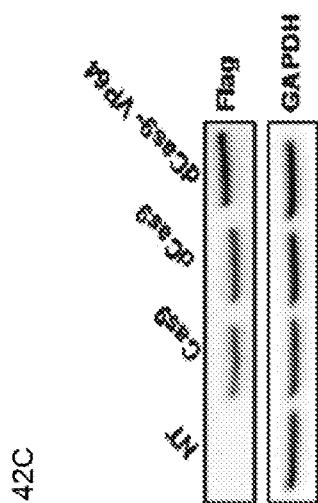


Fig. 42C

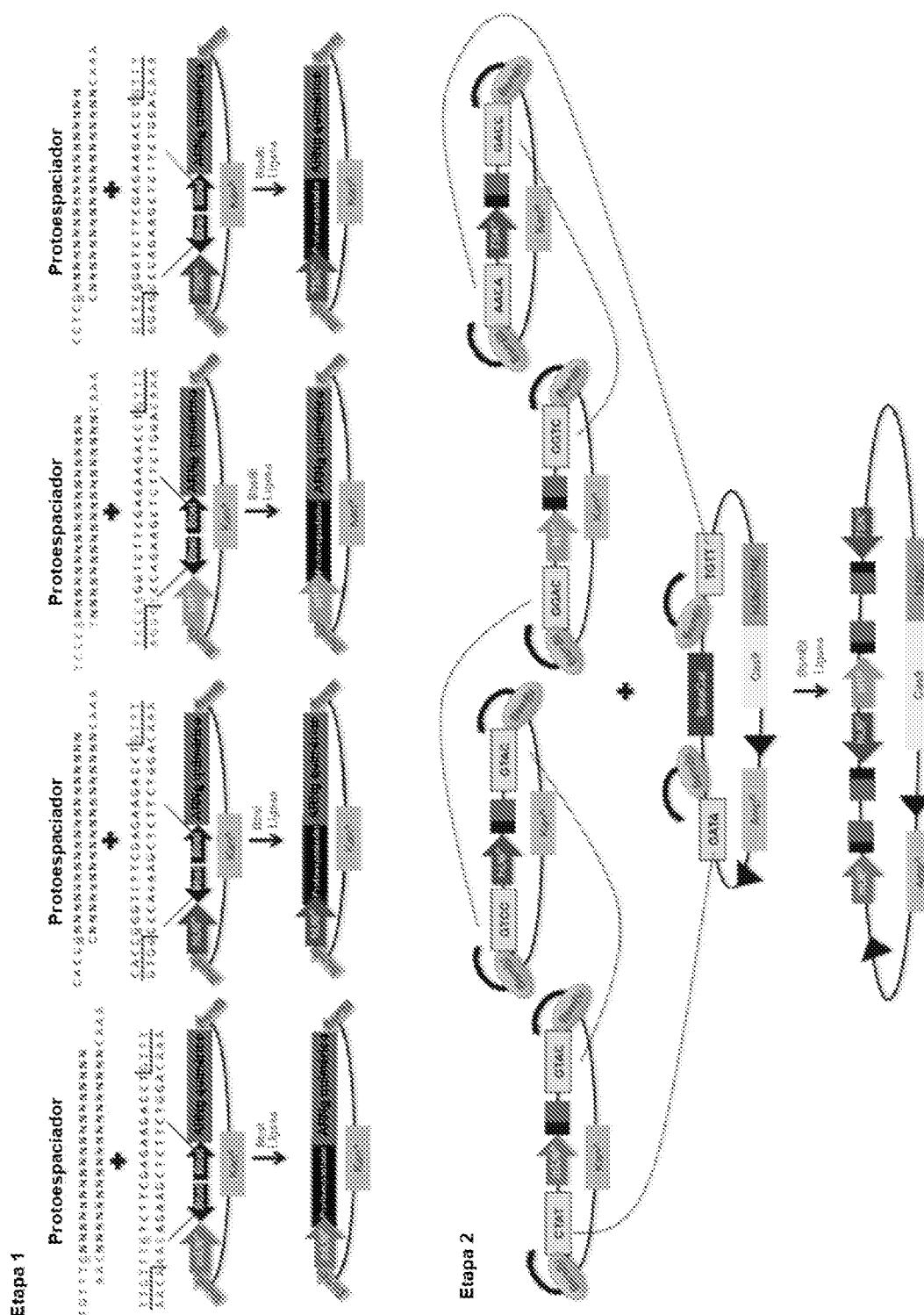


Fig. 43

Fig. 44A

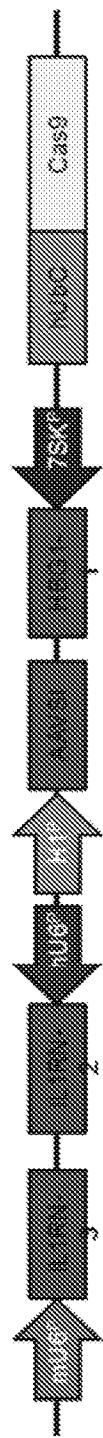
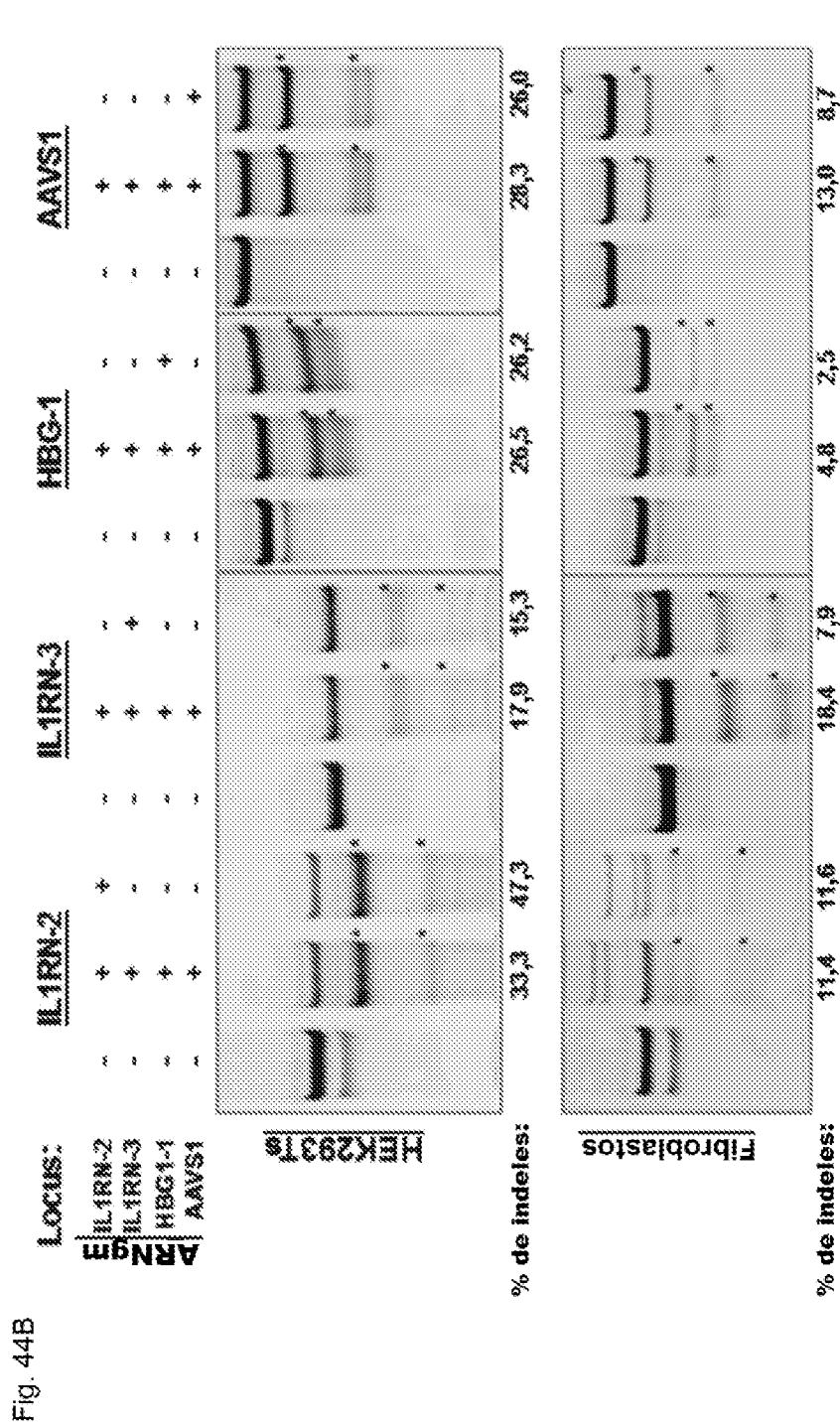


Fig. 44B



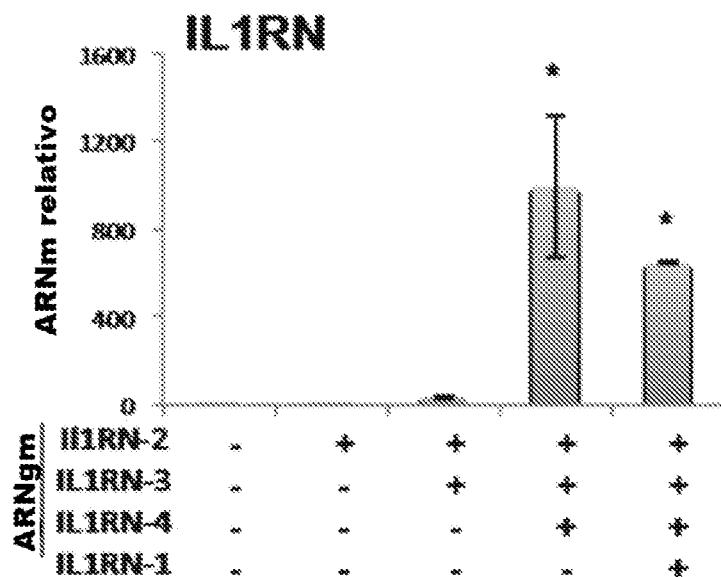


Fig. 45A

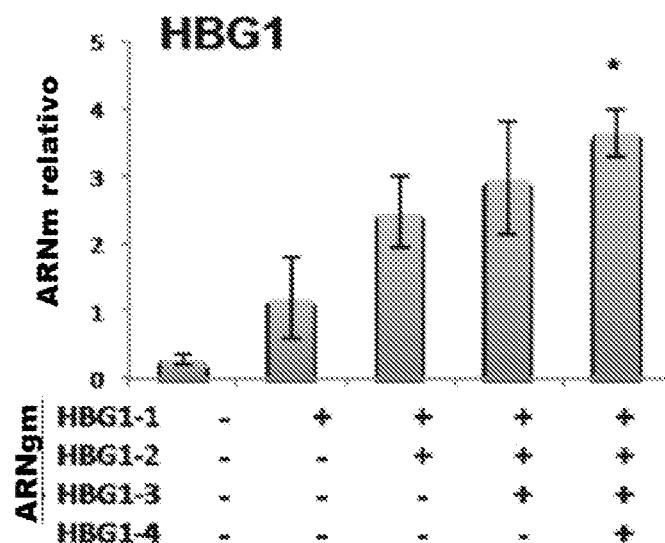


Fig. 45B

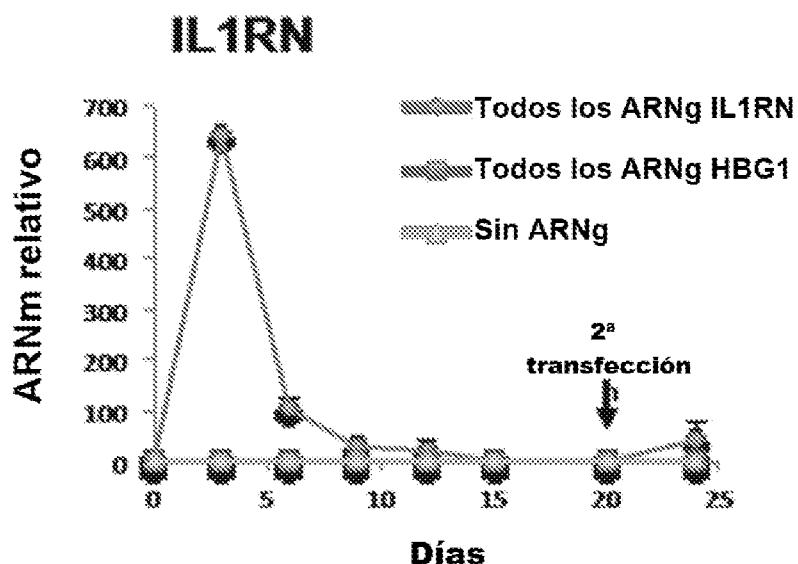


Fig. 45C

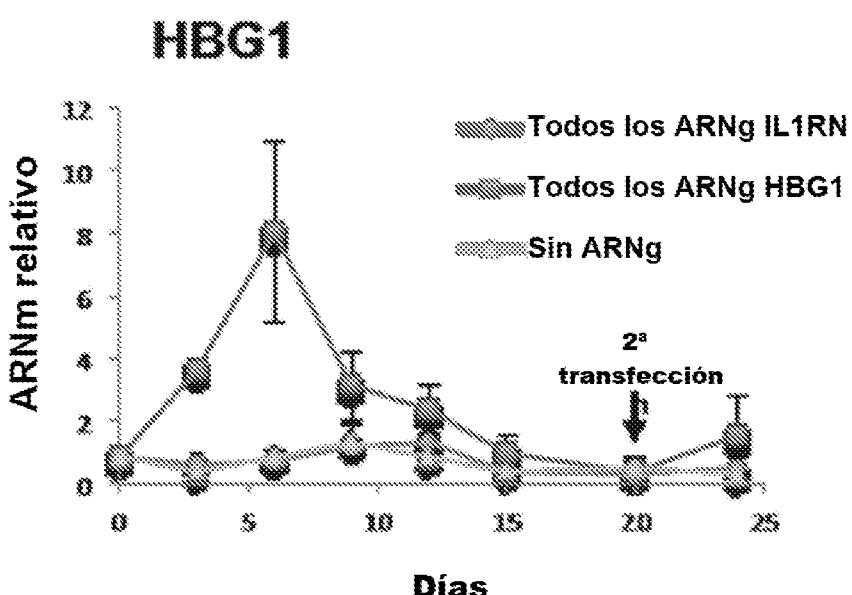


Fig. 45D

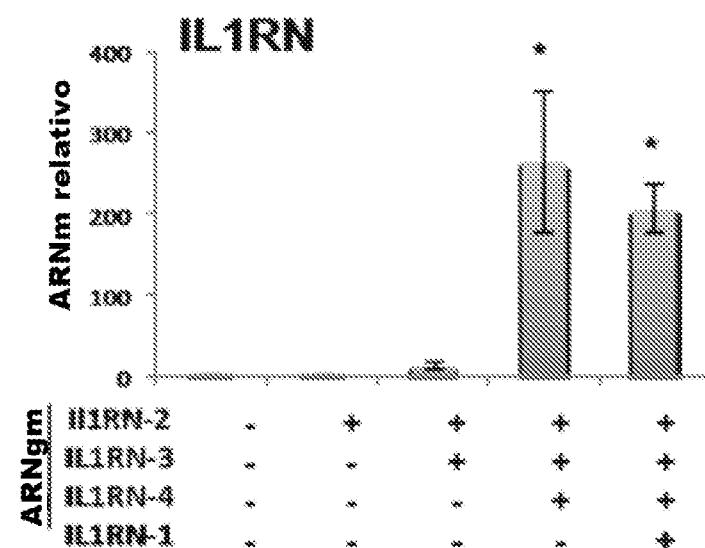


Fig. 46A

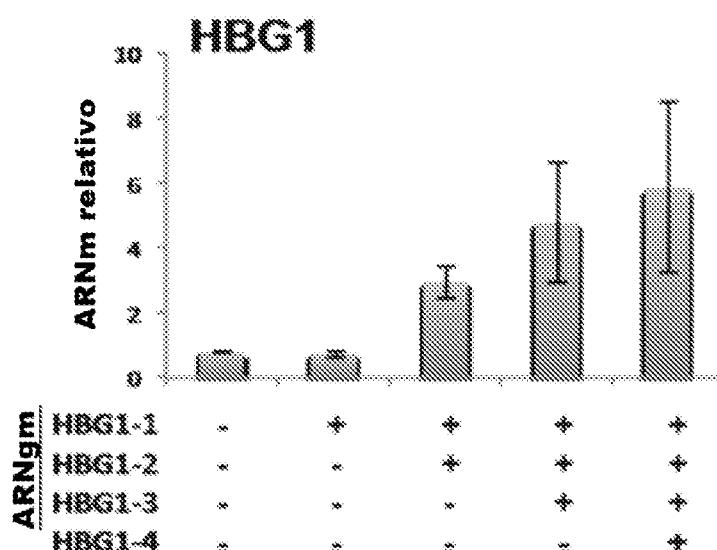


Fig. 46B

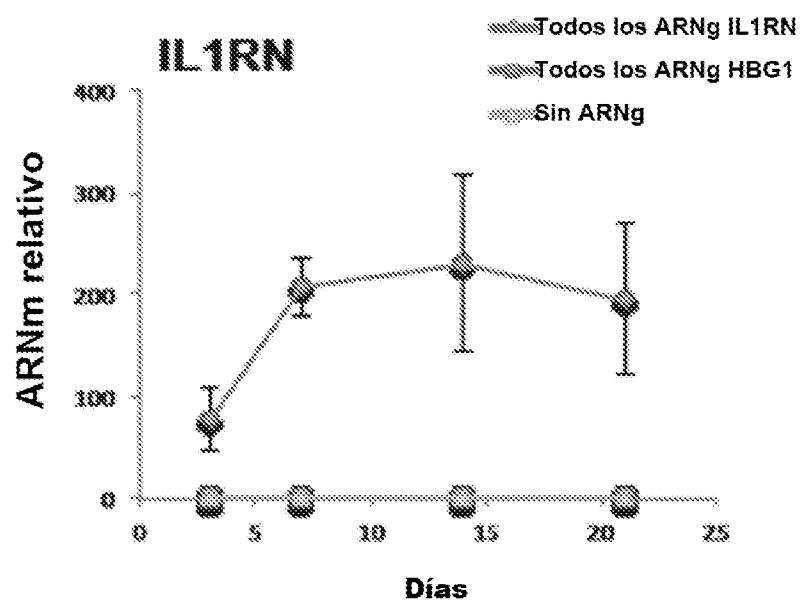


Fig. 46C

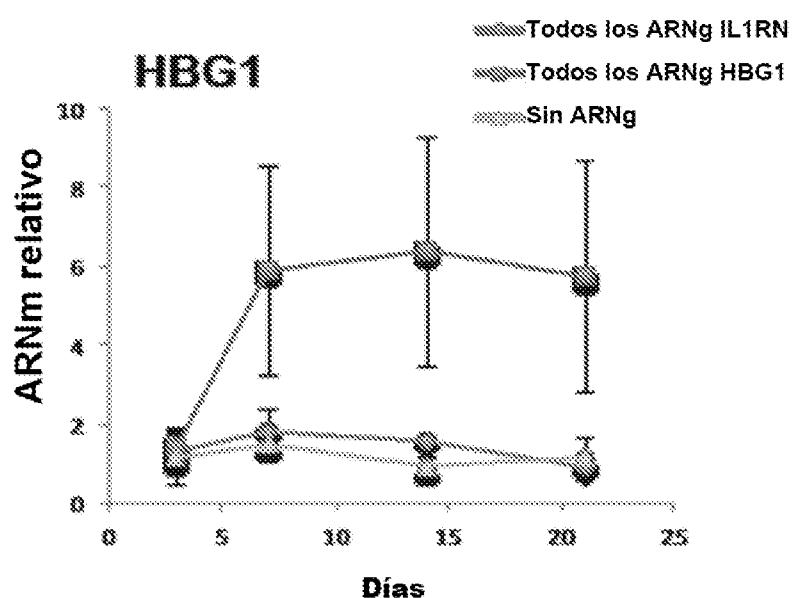


Fig. 46D

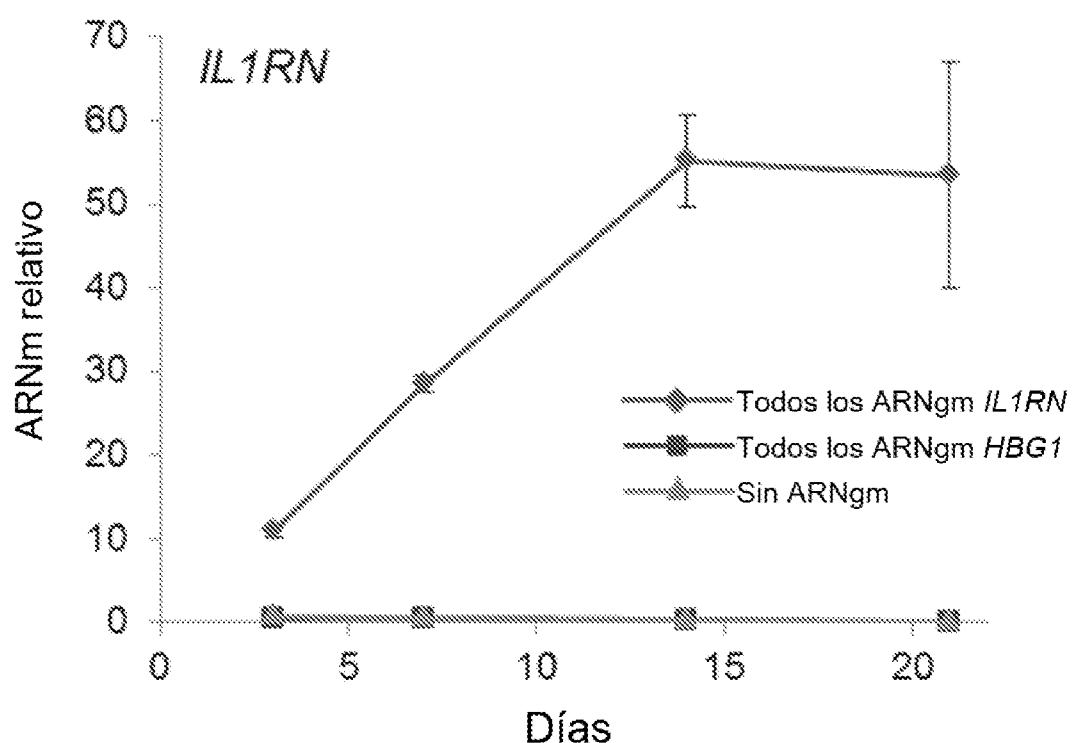


Fig. 47

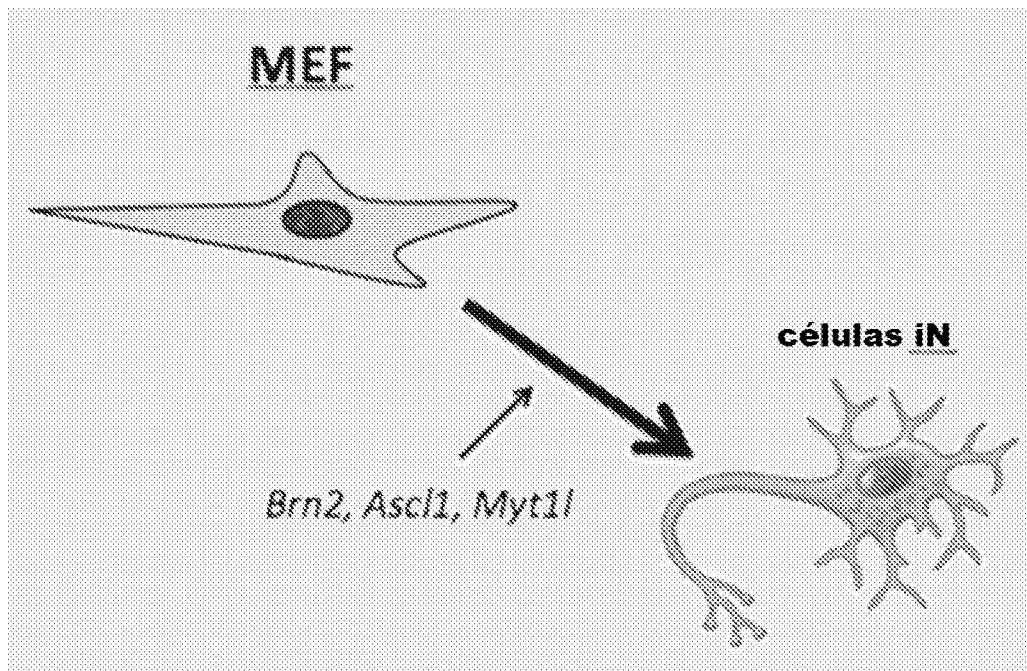


Fig. 48

Fig. 49A

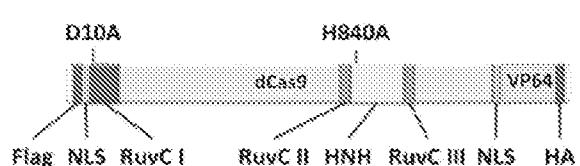


Fig. 49C

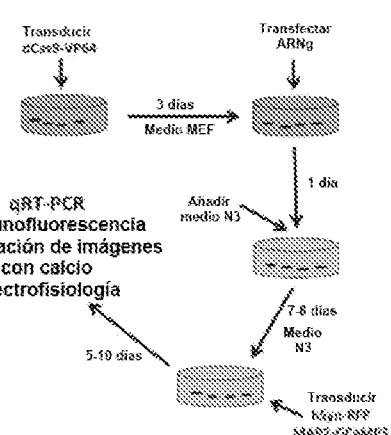


Fig. 49B

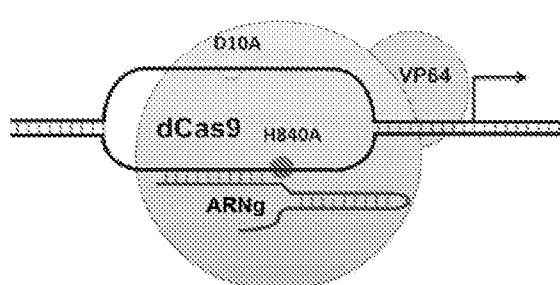


Fig. 50A

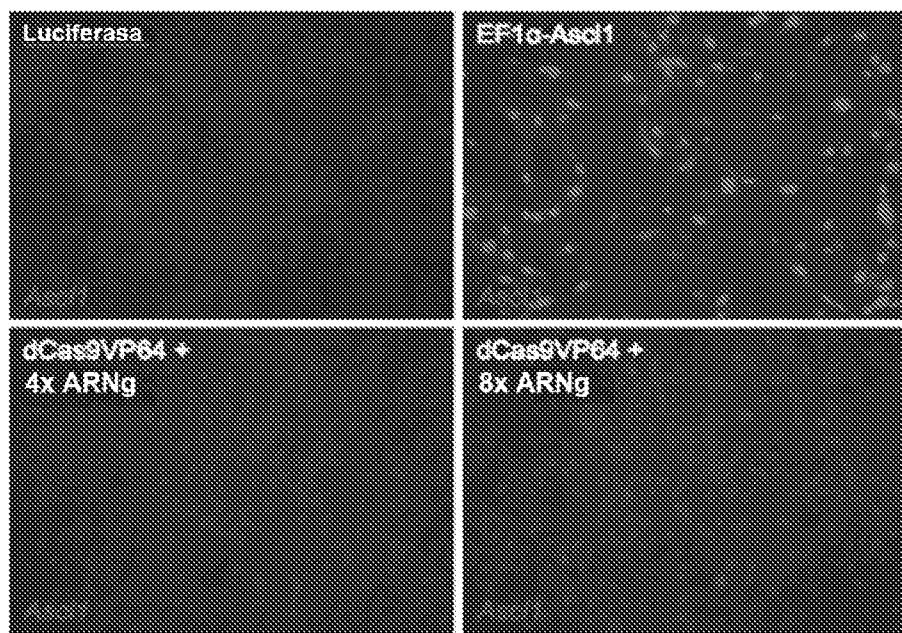
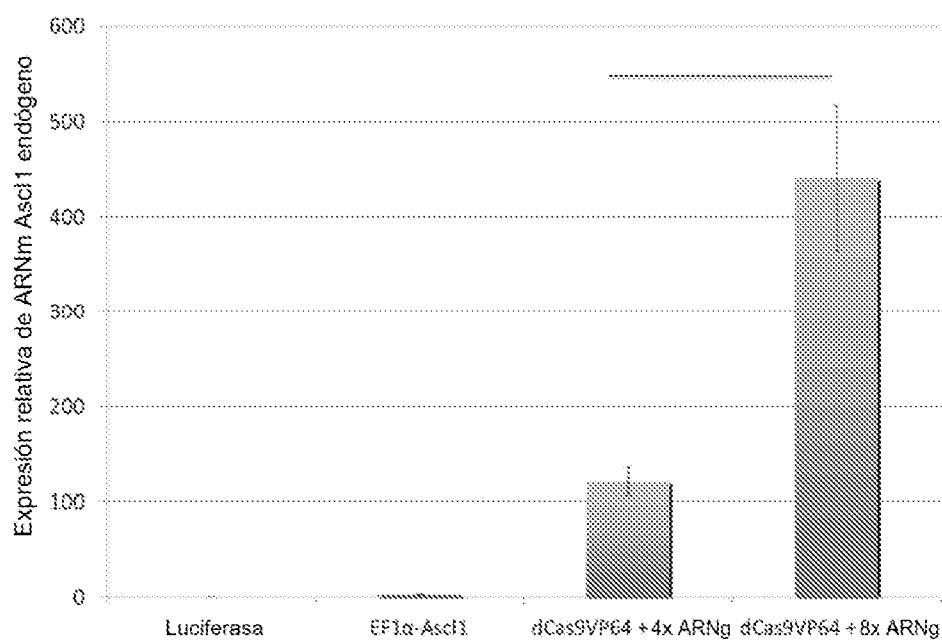


Fig. 50B

ES 2 987 399 T3

Fig. 51A

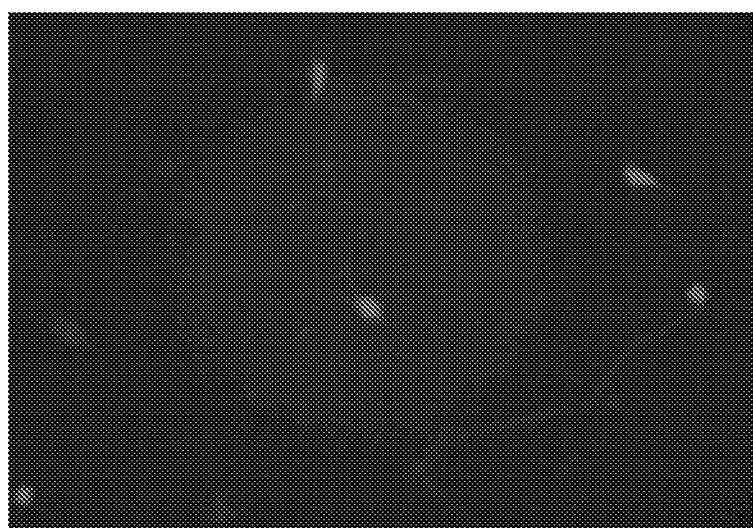
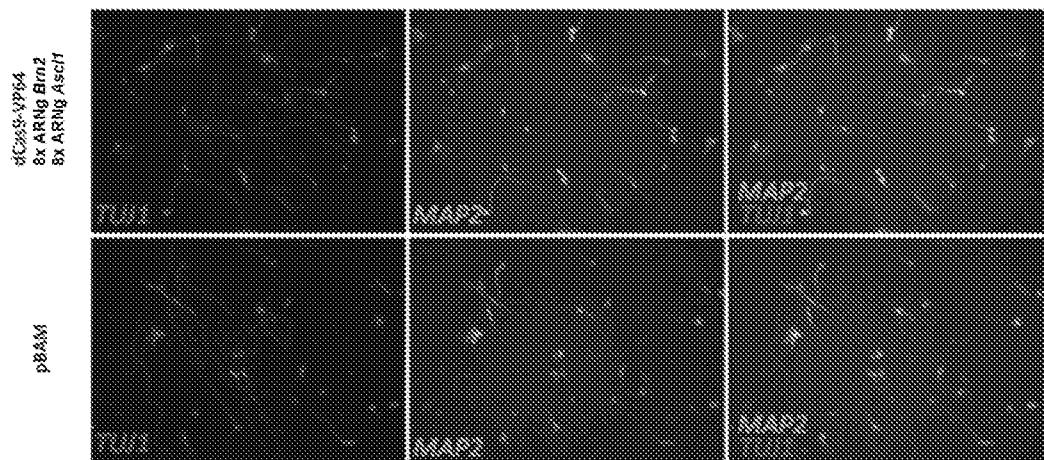


Fig. 51B

ES 2 987 399 T3

Fig. 52A



ES 2 987 399 T3

Fig. 52B

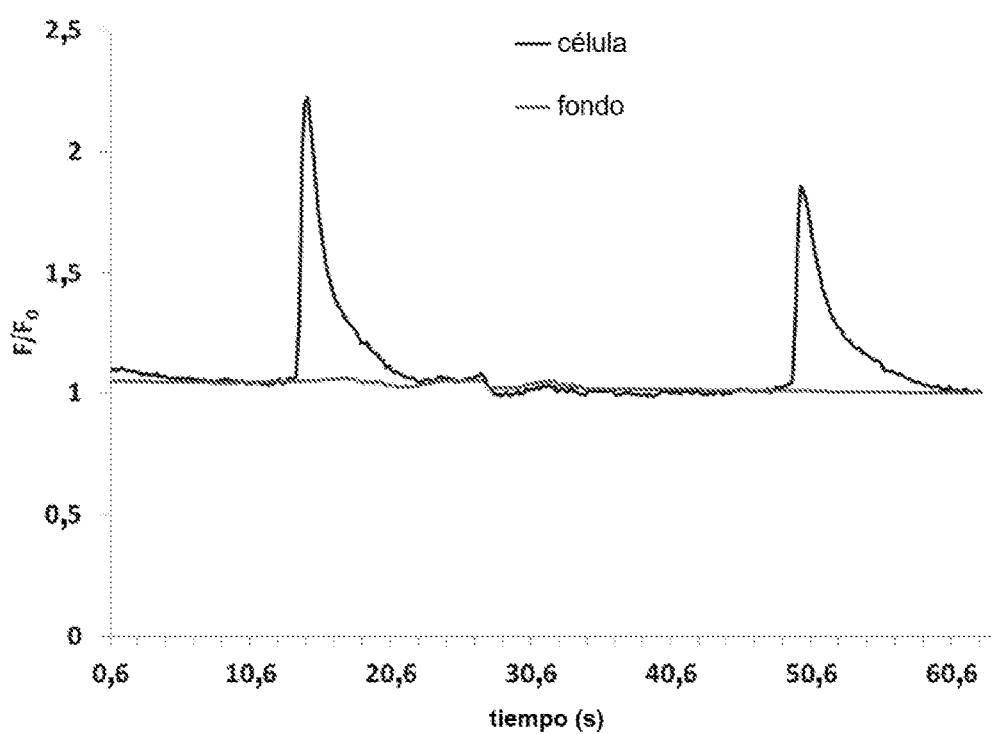


Fig. 53A

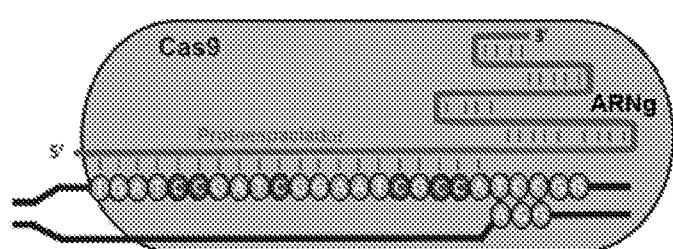


Fig. 53B

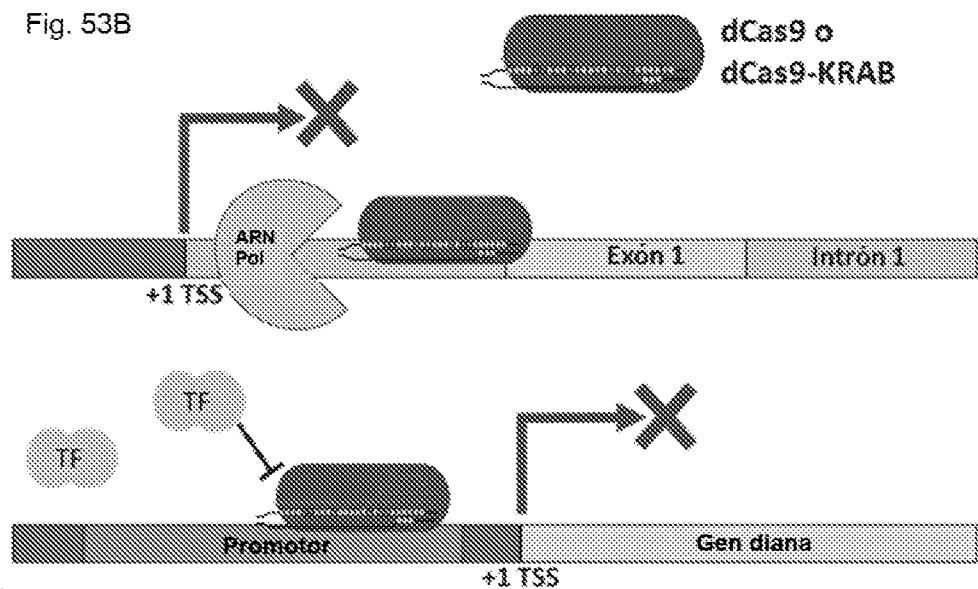


Fig. 53C



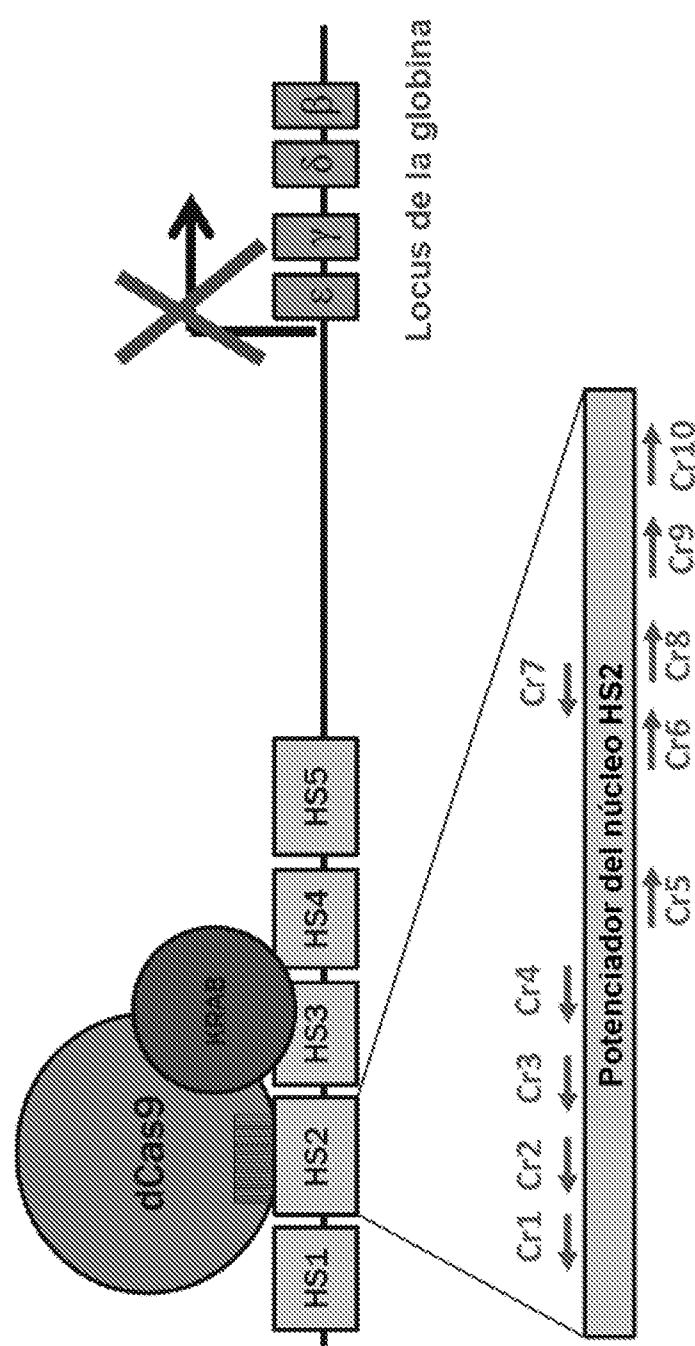


Fig. 54

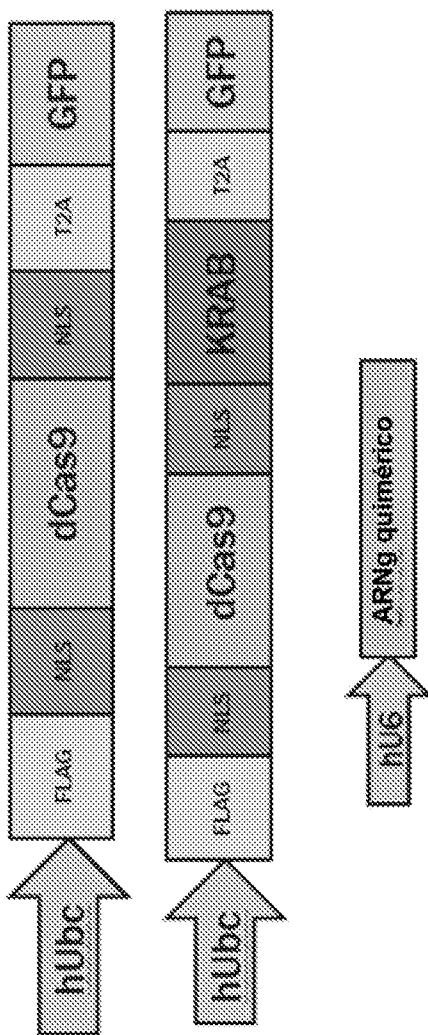


Fig. 55A

Fig. 55B

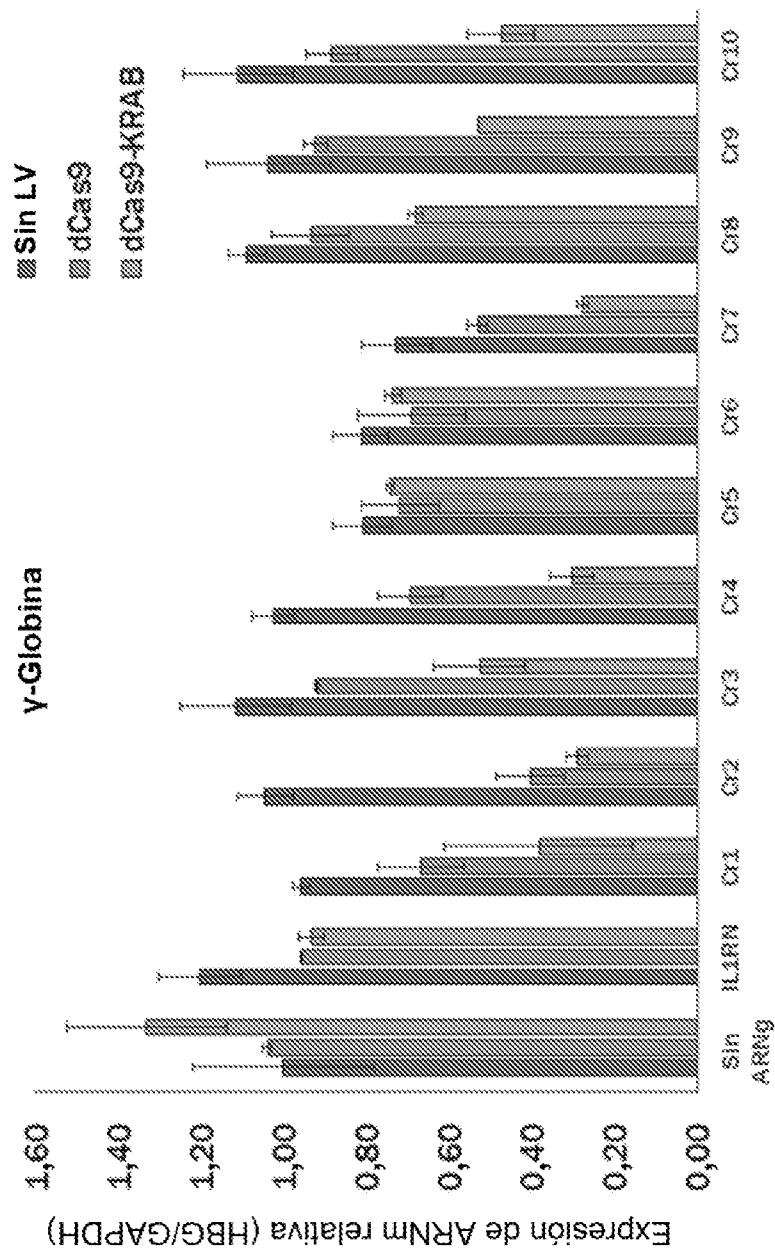
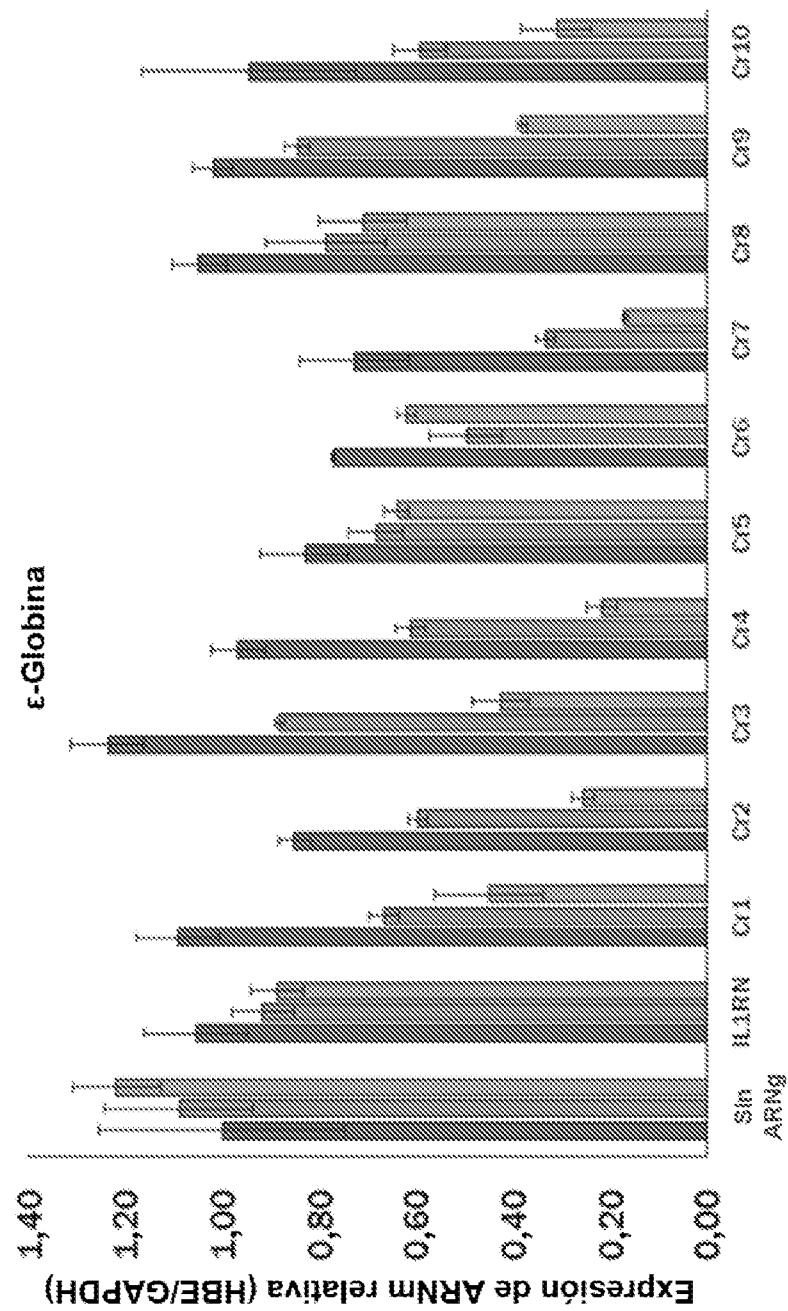


Fig. 55C



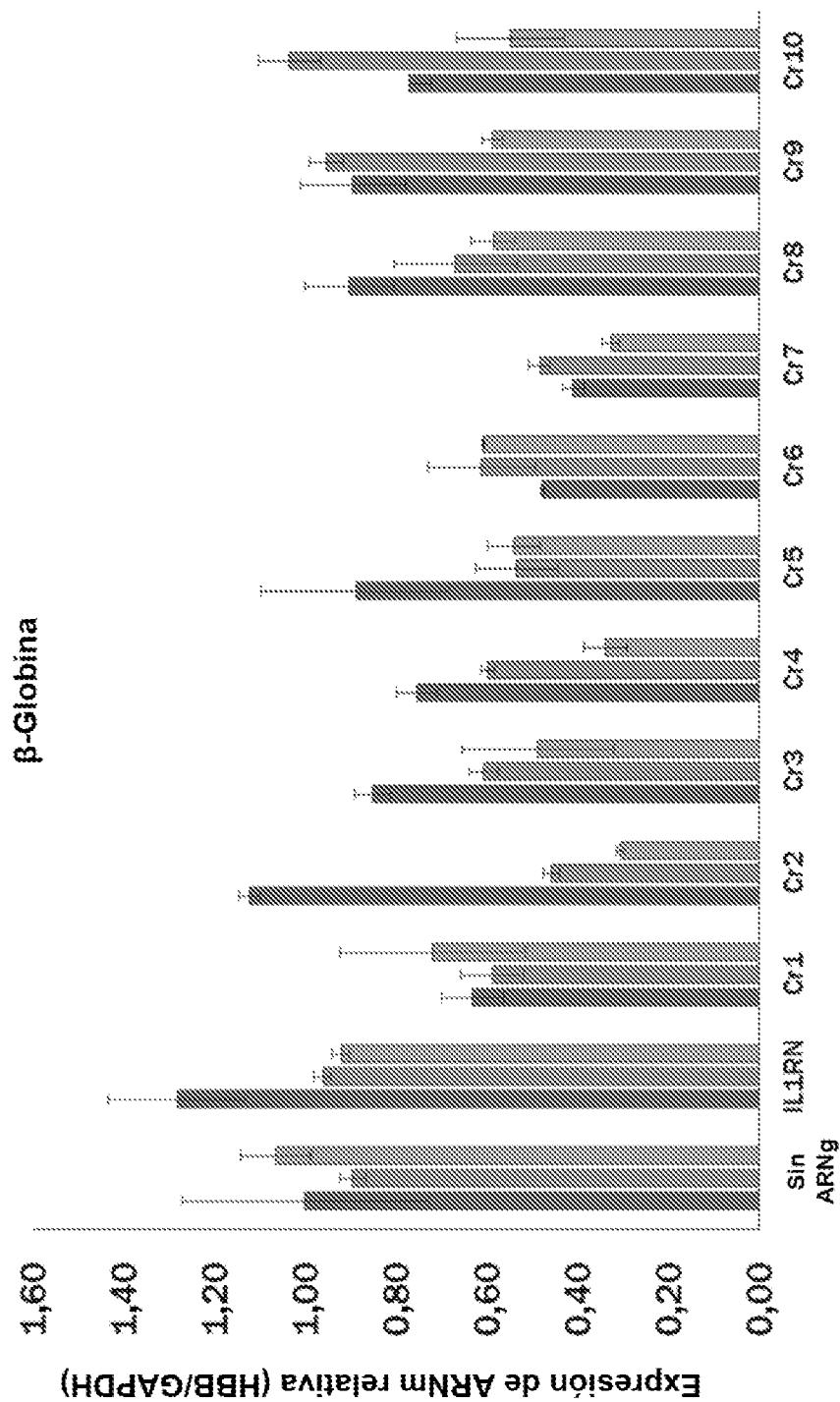


Fig. 55D

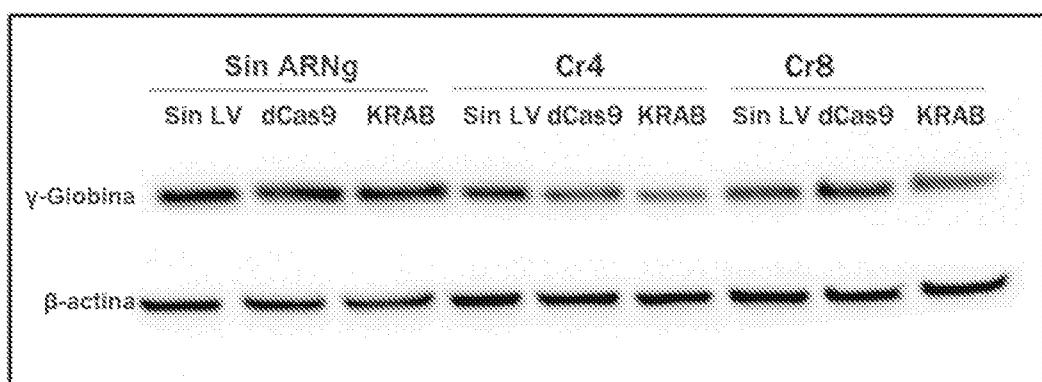


Fig. 55E

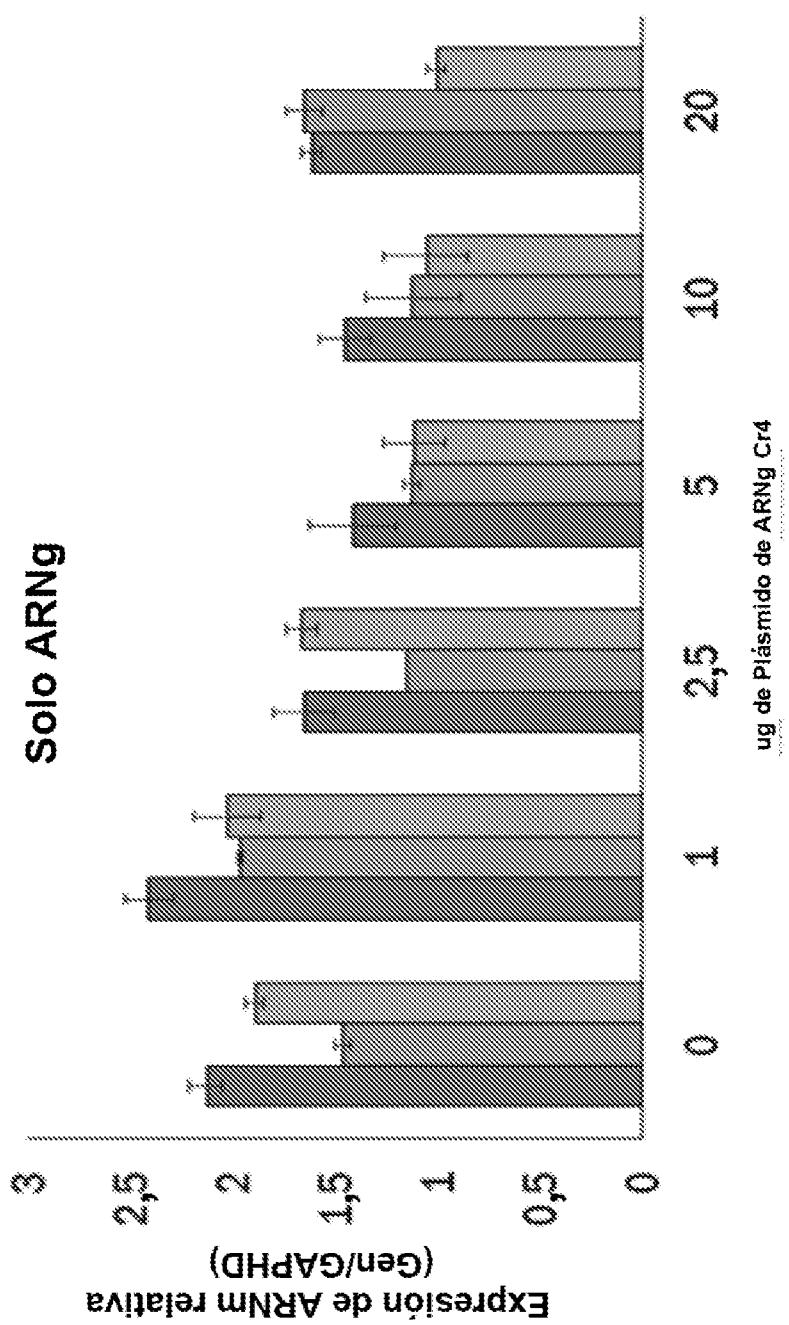


Fig. 56A

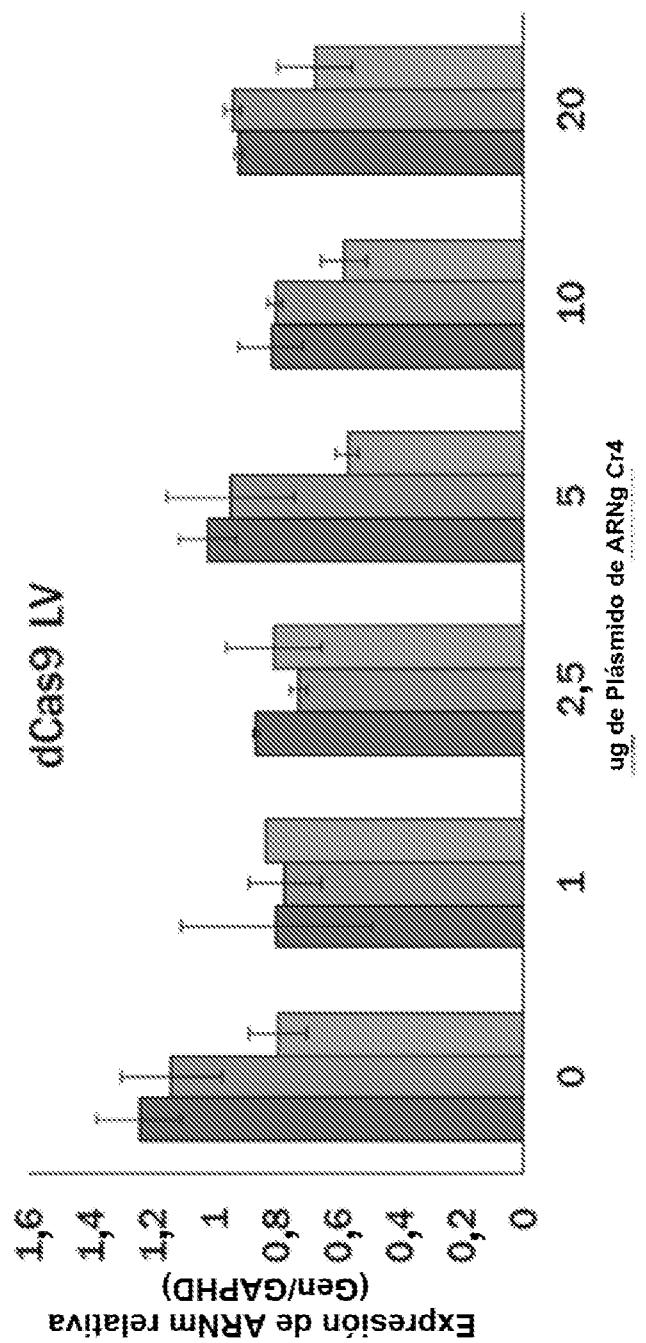


Fig. S6B

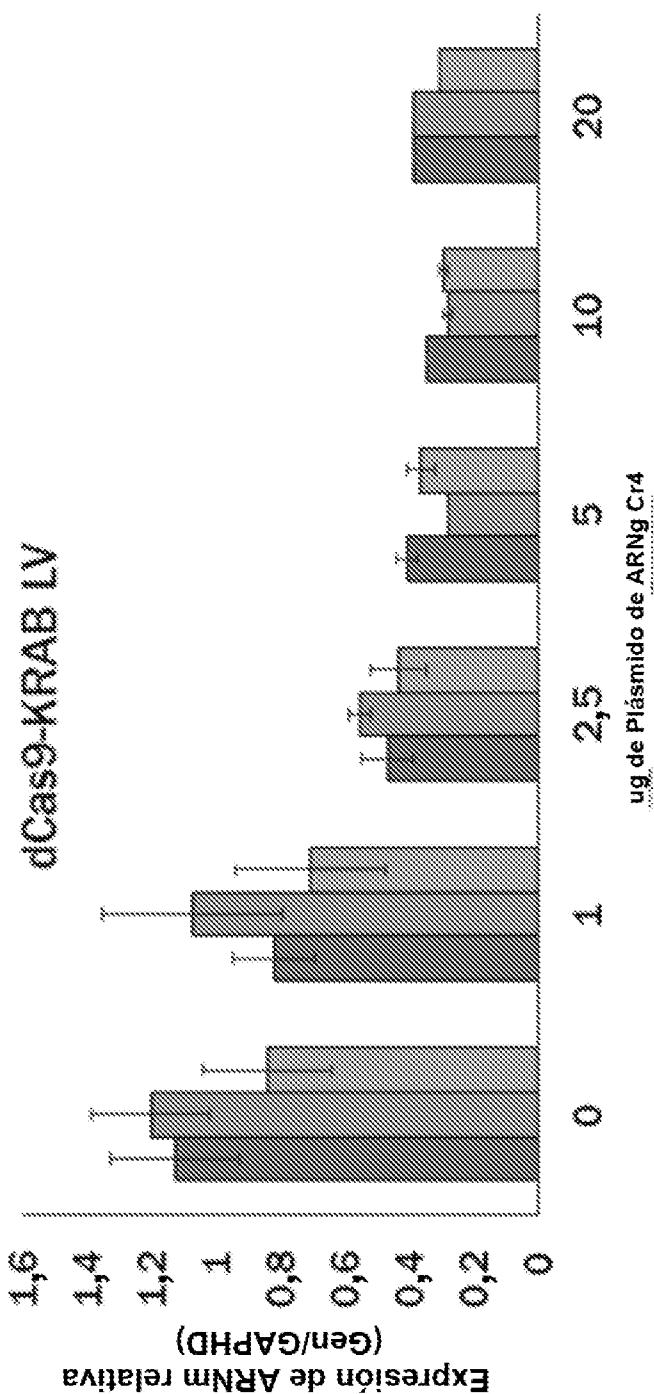


Fig. S6C

Fig. 57A

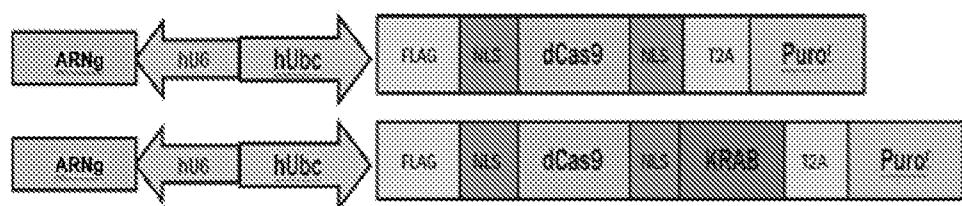


Fig. 57B

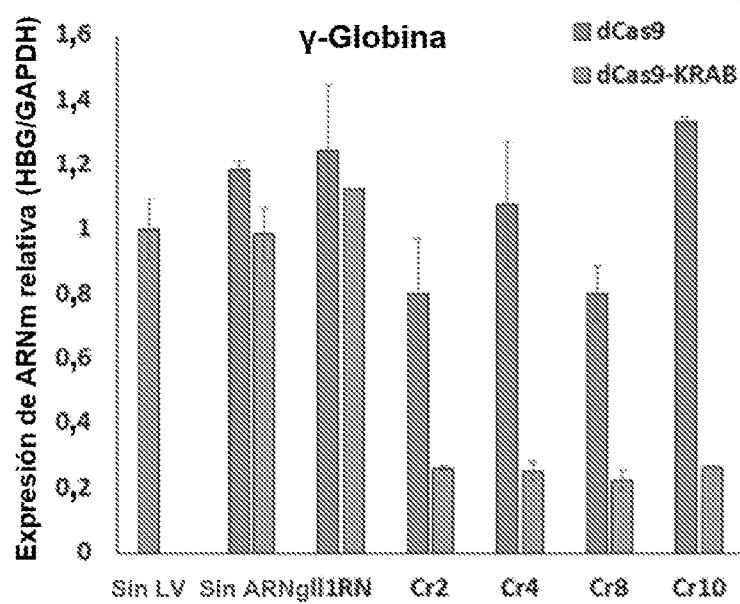


Fig. 57C

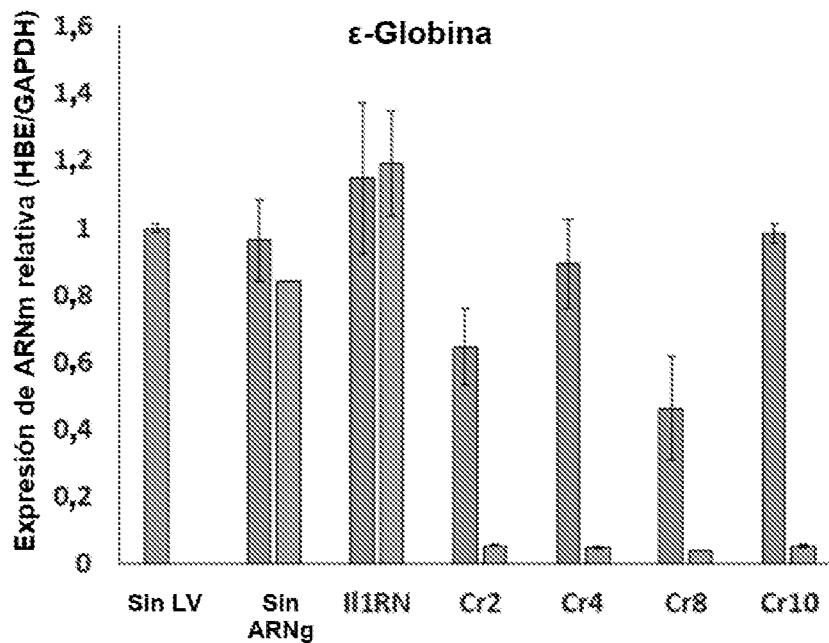
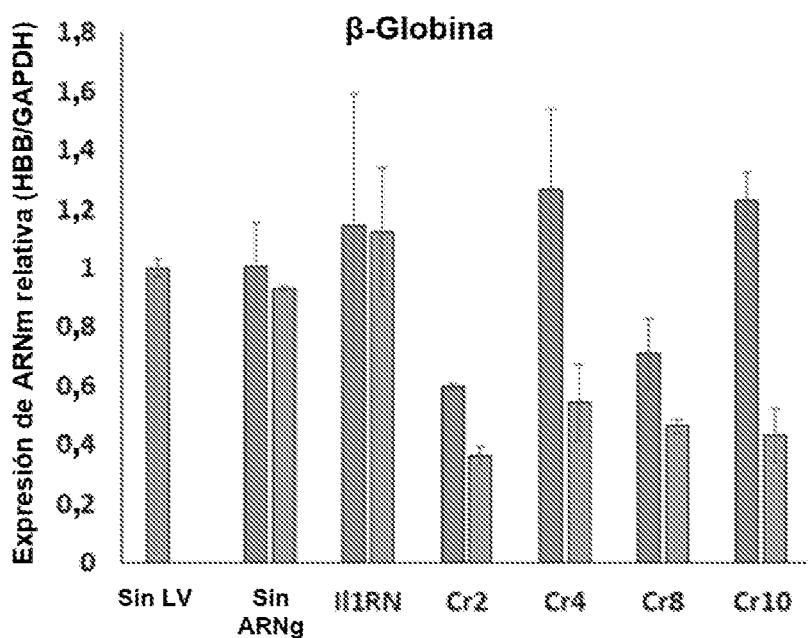


Fig. 57D



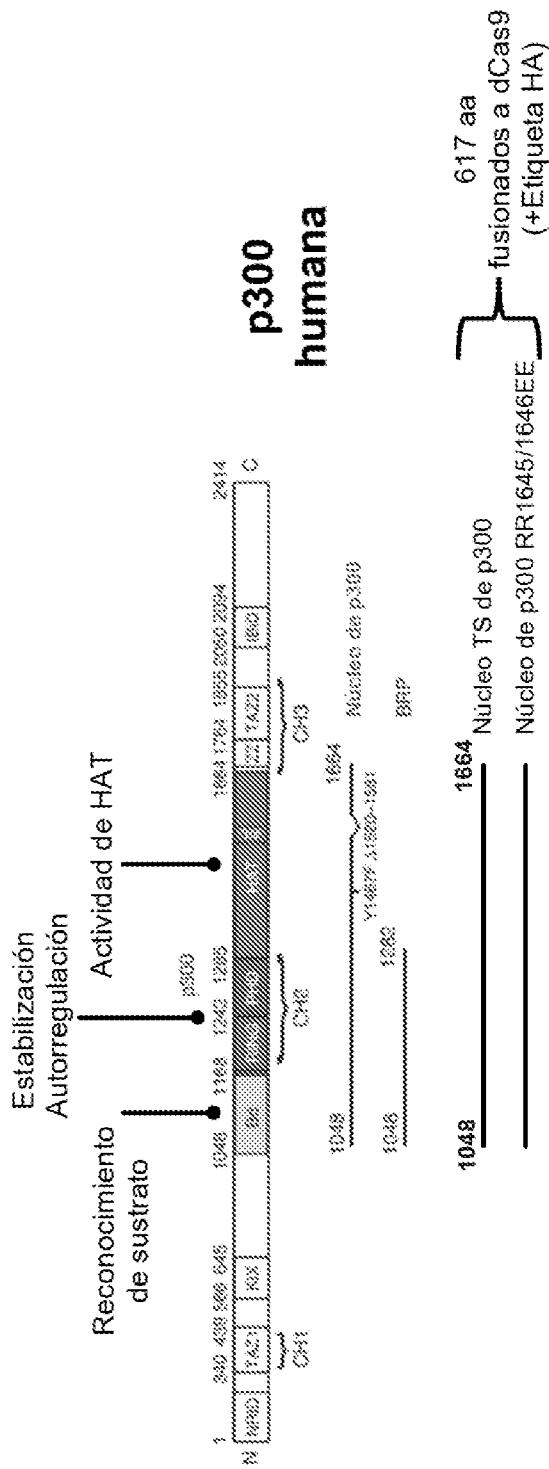


Fig. 58

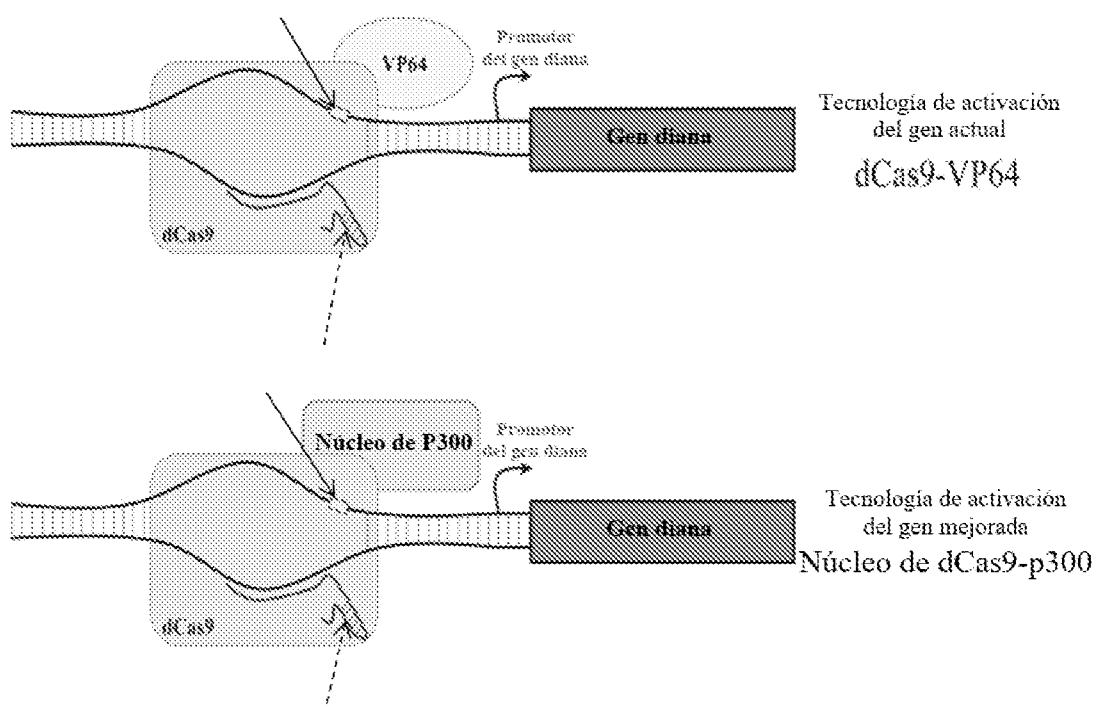


Fig. 59

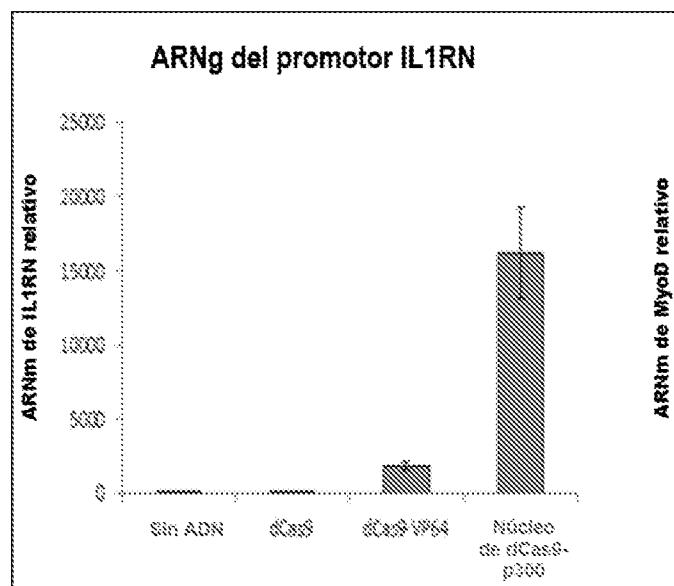


Fig. 60A

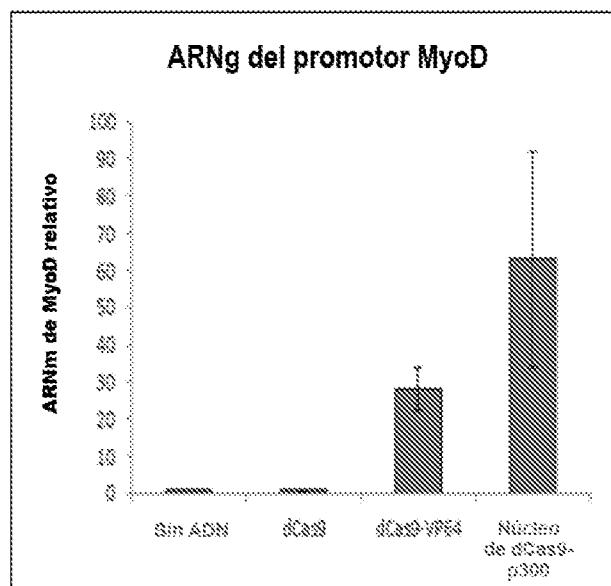


Fig. 60B

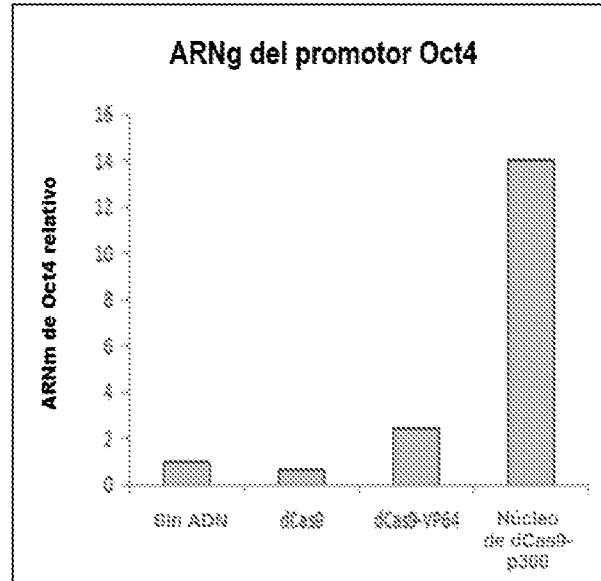


Fig. 60C

i. dCas9

MDYKDHDGDDYKDMD IDYKODDDDMAPKPKRKVGRGNDKYS IGLAIGTNSVGVAVITDEY
 KVPSEKKPKVILGNTDRHSIKKNLICALLFDGETAEATRLKRTARRRITRRKNA ICYLOEI
 FSNEMAKVNDSPFHRLLEESFLVEEDERHHRHPTFGNTIVDEVAYBEEKYPTIYNLSKELVDS
 TEKADILRLIYLALANMIKFRGHFLIEGDLNPONSIYDKLFIQQLVQTYNQLEEEENEINASG
 VDAKAILSARLSKSRRLENLIACLPGEEKNGLPGNLIALSLLGLTPNFKSNFCLAEZAKLO
 LSPDQTYDDLDNLILLAOIGDOYAILPLAAKNLSDATIILSDFILRVNTETITRAPSASMTKRY
 DEHNQDTLLKALVROQLPKSYKEIFFDQSKNGYAGYIDOGASCEEFYKPIKEILEMIDG
 TEELLVKINPEDIILRKCRTFONGSIPHOIHLGELHAIKRCQEOFYFFLKDNRKIEKILT
 FRIPYVVGFLARGNSKFAASMTRESEETITTPWNFFFEVVDEKGASAQSFIERMTRNFIRKLPNE
 XVLPEKHLLYEYFTVYNNELTKVVKYTTEGMNRPAELSGEORKEATIVDILLSECTNEEVTVRQLK
 EDYFKKIICFDSVEISGVEIDRFNASLGTYNDLLKIIKDKDFLNEENEDILEDIVLTTL
 FEDREMIIEERLSTYANLFDUDMVNKQLERRRTYTGELSRKLINGIDRQCSGKTILDPLRS
 DGFANENFMCLTHEDSLTFKEDIQSAQVSGCGDLSIHEHIANLAGSPAIKKGILQTVRYVD
 ELIVKVMGRHXDPEXIYIEMARENQITQNGQFNSRERMKRIEEGIFELGSQILKRNHPVENTQ
 LONEKILYLYYLONGRDIMYDQEELIDINELSOYIDVDRIVPQSFLEKDDSIDMKVLTSDRNREG
 KSDINVRPSEEVYKEMNNYWRQLLNAKLITQKFIDNLTKAERGGSELIXAGFIFRQLVETR
 QITKHVAQILDSRMNTKYDENOKLIREVKVITLKSFLVSDFRKDPQFYKVREINYYINHAH
 DAYLNAVVGTLAIKEYPRLESEFPVYGDYKVIIVRRENIAKSEOEIGKATAKYFFYBNINMF
 EKTEITLANGEIIEKRPLIFTNGETGEIVWIKGRDFATVVKVLSMPQVNIVKETEVOTGGF
 SKESIILPKRNSDELIAKRNWMDPKYGGFDSEPTVAYSVLVVAKVKGGSKKLLKSYKELLG
 ITTMRERSSEKNPIDFLFANGYMEVVKDILITIPLPEYSLFELNGCRKRMJASACKLORGNE
 LALESKYVNNFLYLAQHYEFLKGSPEDNEQFOLEFVEQHKHYLDEIIEQISKEFSERVILADA
 NLDXVILSAYNMRDREPIREQAEINI IHLFTLTNLGAPAAFKYFOTTIDRERRYTSTKEVLD
 TLJHQSIITGLYETRIDLSQLCGGIPKKKKRKVG

Leyenda de aminoácidos:

Epítopo "Flag"

Secuencia de localización nuclear

Streptococcus pyogenes Cas9 (D10A, R840A)

Efector VP64

Efecto del núcleo de p300

Epítopo "HA"

Fig. 61A

ii. α Cas 9^{W⁴⁴}

MUYKONEDGODYEDNDIJOYKDDDDOMRMPKKKVKGRGMDKKYSIGLAIIGENSYEWAYITDEY
 KVPSSKKFKVILGNTDRHSIKMNLIGALLFDSGETAAEATLKRTRRRTTERRKNRICYLOEI
 FSNEMARVDDSFHRIEESFLVEEIKEHEBHPIFGNIVDEVAYHEKYPTIYHLRFKLVDST
 TDRAOLSLIYLALAHNIKFEGHFLIEGDLNPINSNDIKLFICLVCOTYNQLFEENPINASG
 VDAAKAILSARLSKSRRLENLIACQLPGEKNGLFGNLLIALSILGLTPNFKSNFDLAEDAKLQ
 LSKDTYDUDLDNLIAOIGDCYADOLFBLAANLSDAILLSDILRVNTEITPAAPLSASMIERY
 DEHHQDLTLIKALUVRGQLEPKYKEIFFDQSKNGYAGYIDEGASOEEFYKEIKSIILESMING
 TEELLVNLNREDLLRFQRTFDNGSIPHGICILGELHAILNRCQDFYPFLDNREFEXILIT
 FRIPIYYVGPLARGNSRFAXXMTRESEETITPEWSFFEVVDEKGASAOSFIERMTNFURNLPNE
 KVLEKHSILLIEYFTVYNELTIVKVKYVTEGMKXKAFLSGEQKAIVDLLEKTNRKVTVVKQIK
 EDYFKKIECFDSVEISGVEDRPNASLGTYHDLLKIKRDKDFFDNEENEDILEDIVLTL
 FEDREMEIERLKTYYAHLFOOKVYKQLEKBRBRRTGWERLRLKINGIRDKOSGKTIILDPLNS
 DGFANENEMOLIHDDSLTFKEDICOKAOVSQGDSLHEHTIANLASSPATKEEGILQTVKVD
 ELYKVYNGRHKPENIVIENARENQTTQXGGQXSRERMKIEEGIKELGSQLILKENFVENTQ
 LONEKLYLYLONGLRDMYDQELDINRLSDYDVDAIVPOGFIKDDSIDNNVILTRSDKNRG
 KSINIVPSFEVVVKMMENTYWRCLLNAKLITQREKFDNLTKAERGGLSELDKAGFIEROLVETR
 QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNNYHHA
 DAYILNAVYGTALIKKYFKLESKFVYGOYKVKYDVRKGMIAKSEOEIGKATAKYFFYSNIMNF
 FKTEITLANGEIRKRELIELTNGETGEIVWDKGRDFATVKEVLSMPGCVNIVKRTTEVQTGGF
 SKESITLFRMSDELJARKKNDPERRYGGFDSPTVAYSVLNVAKVERGKSEKLLSVKELLG
 ITIMERSSPFMNFIDSTLEANGYKEVKDILIIKLPKYSILFELENGEKPMLASAGEILQKGNE
 LAEPSEYVNFLYLAŞEYEKLEGSPEDNEONQOLFVEOHEHYLDEIIEPOISEFSKEVILADA
 NLCKVILSAYNEHEDKIIREQAEINIİHLFTLTNLGAPALARSKYFDTTIDRKEYTSTKEVILADA
 TLİHQSITGLEYETRİDLSCLOGDFIAGSKASPKKKRMVGRADALDDFDLDMLGSDALDDF
 DLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDF

Fig. 61B

¶ i. dCas9^{Nucleo de pM9}

M.DYKNDHDGDIKDMDIDYKDDDDMAPXXXKRNKGGRGMKRYSIGLAFGINSVGWAVITDEY
 KVPSEKEFKVLGNTDRHSIEKNLICALLFDGETAATRLEPTAERRYTRBKNSHICYLOEI
 FSNEMAKVDDSFHHLLEESFLVEEIKERHPTIFGNIVDEVAYHEKETIYHILRKLVDSD
 TDKAIDLALIYLALASMIKFRGHFLIEGDLNPINSVDKLFICQLVQTYNQLFEENFINASC
 VDAAKAILSLSKSFRLENLIAOLPGERKNGLFGNLLIALSILGLTPNFKSNFOLAKDARLO
 LSKDTYDODLDNLILAQIGDGQADLFLAAKNLSDAIIISDILSYNTETKAPLSASMIKEY
 DEHHQDLTLLKALVPROOLPEKYREIFFDOSKNCYAGYIDGGASCEEEFYRFIPLKMDG
 TEELLVKLNREDLILRHOFTFNGSIEFHOTHLCELNATLRFQEDFTYFFLFDNREKIEKILT
 FPIPYVVGPALARNSSFANMTERESEEITITTEWNFEEVYDKGASAQSFIERNMTNFKNLENE
 KVLPKRSLLVEYPTVYNELTFKVVKVTECMRKPAFLSGEOXKATVLLSKTNRNVTVYQLK
 EDYFPKEIECFDSVIEISCVEDRPNASLGTYHDILLKIIKOKDUSLONEENEDILEDIVLTLL
 FEDREMIEEBLKTYBMLFDOKVMQOLKRRBRYTGNGRILSRKLINGIRDKQSGKTILDFLES
 DGFANRNFNQOLIMDOSLTFFREDIQXAOVSOGGDELMEHTANLAGSPATENGILQTYKVD
 ELVKVNMRHKPENIVIEMAFENQTTORGQNSRERMKRIEEGIFELGSQILMEHPVENTO
 LONEKLYLYLONGLIMIVDQELDINFLSDYIVDATVPOSFLDDSDDNKVILTRSDENRG
 KSINVPSEEVVKEMNIVWEGQLNAAKLTQRKFDNLTKAERGGLSELDRKAGFIRSQLVETR
 QITKHVAQILDGRMNTKYDENDKLIREVKVITLKSFLVSDFRKDFQFYRVREINNNYRHAAH
 DAYLNRYVGTALIKSYPELESEFVYGEYKYYDVREMIAKSEFOIGKATAKYFFFYSNIMNF
 EXTEITLANGEIRKRPILLETNGETGEIVWDXGRDFATVRKVLSMPQVNIVKETEVQTCGGF
 SNEISILPKRNSDKLIAKRDIDERKIGGFDSPTVAYSVLVVAKVEKGSKKLKSVEELLG
 ITIMERSSSEKENPICFLEANGYEEVREKILLIILPKYSLFELENGKRMLASAGELQEGNE
 LALESKEYVNFLYLAZYEKLIKGGPDENEQEOLEVECHKHYLDEIIROIISEFSERVILADA
 NLDKVLISAYNKHRONPIREQAFENIINLFTLTNLGAPAAFKYFDTTIDENRYTSTKVEILDA
 TLIHQSIITGLYETRIDLSQLCGDFIAGSKASPKKKKRVGRAIFKPEELRQALMPTLEALY
 RODPESLPLFROPVDFOLLGIPDYFDIVKSPMILSTIKRKLIQTOGOYQEPWQYVDDIWLMFN
 NAWLYNRKTSRVYXYCSNLSEVFSQIDPVMOSLGYCCGRKQLEFSPOTLCCYGNQLCITP
 RDATYYSYONRYHFCEKCFNEIOGESVSLCOOPSOPOTTINKEOFSKRMNDTISPELEFE
 CTECGGRKMHQICVLUHELIIWPGFVCDUCLAKSARTKRENKFSAKRLPSTRIGTFLENRY
 NDFLRBNHPESEGVTVRVVNAEOKTYEVKRCMKAEFVDSGMAESSEPIKTKLFAFFEL
 DGVOLCFPMHIVCEYGSUOFPNORRVYIISYLDSTVHFFYRPKCLRTAVYHEILLIGYLEYVK
 KLGYYTGHRIWACPPSERGOUYIIFHCPRPQOKLPKPKRLOEWYKOMLDKAVSERIVBDYKDI
 EYQRTEDELTSAKELYEEGDFWENVLEESIKELEGEFFFFEEREEETNSNESTIVTKGDSK
 NAXXXNNKNTSKNKSLSRCNXXPGMPNVSNOLSOQLYATMEKHKEYFFVIRLJAGPAA
 NSLPPIVDEDPLIPCOLMDCDAFITLIABKHLFSSLBRAOWSTMOMIVLHTOSODIP
 YIVPDIAS

Fig. 61C

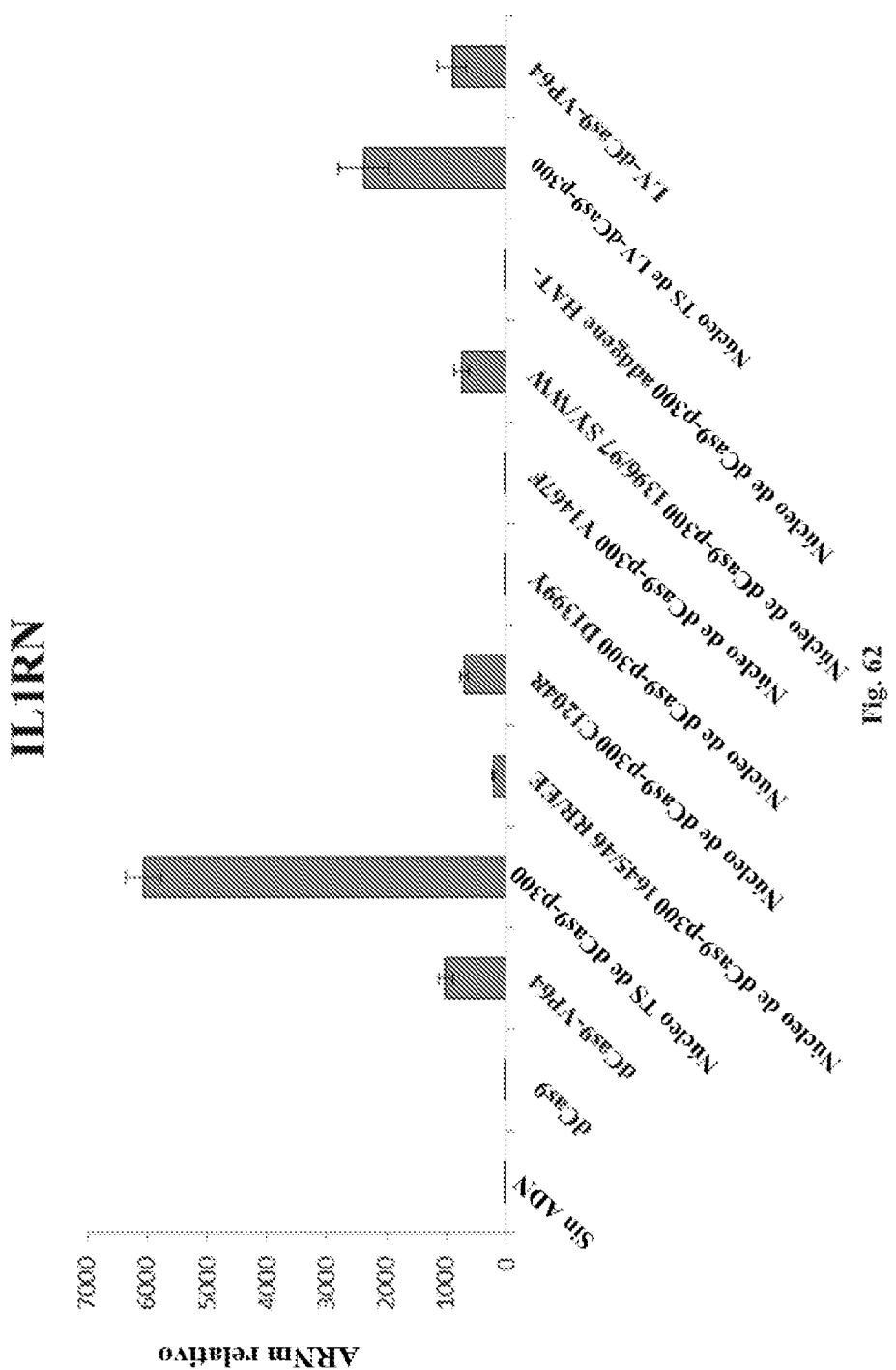


Fig. 63A

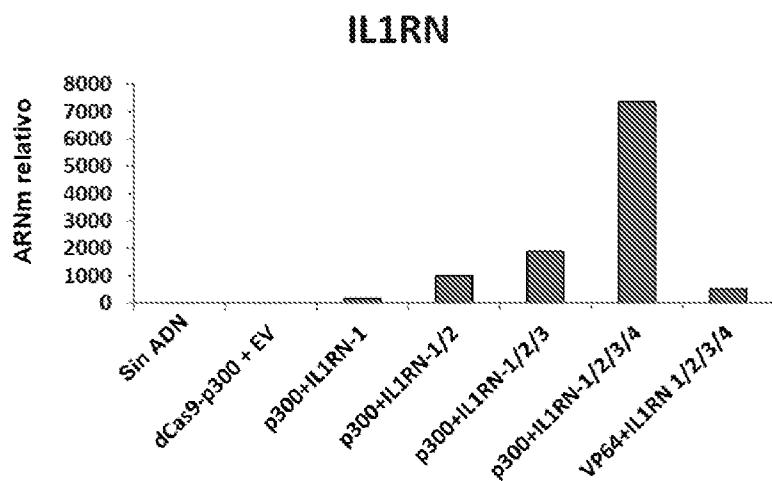


Fig. 63B

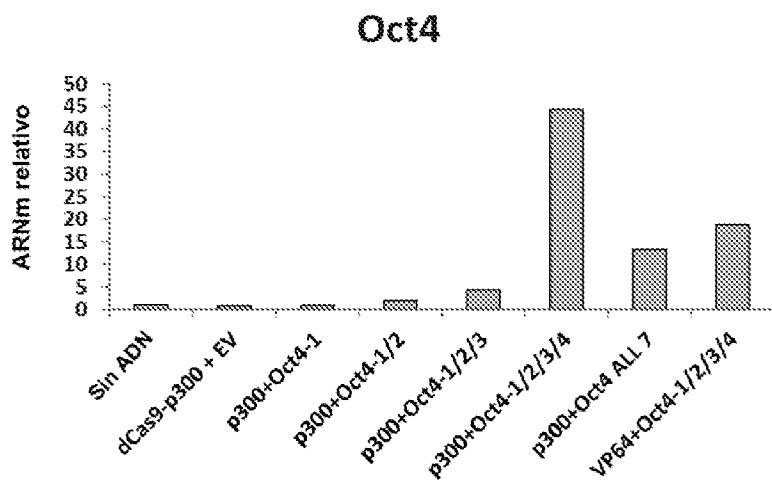


Fig. 64A
PCR genómica para detectar la integración dirigida

