

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 754 433**

51 Int. Cl.:

C12N 9/78 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.12.2014 PCT/US2014/070038**

87 Fecha y número de publicación internacional: **18.06.2015 WO15089406**

96 Fecha de presentación y número de la solicitud europea: **12.12.2014 E 14825518 (5)**

97 Fecha y número de publicación de la concesión europea: **07.08.2019 EP 3080265**

54 Título: **Variantes de Cas para edición génica**

30 Prioridad:

12.12.2013 US 201361915386 P
16.04.2014 US 201461980333 P
08.07.2014 US 201414326318
08.07.2014 US 201414326140
08.07.2014 US 201414326303
08.07.2014 US 201414326290
08.07.2014 US 201414326269
08.07.2014 US 201414326109
08.07.2014 US 201414325815

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
17.04.2020

73 Titular/es:

PRESIDENT AND FELLOWS OF HARVARD COLLEGE (100.0%)
17 Quincy Street
Cambridge, MA 02138, US

72 Inventor/es:

LIU, DAVID, R. y
KOMOR, ALEXIS, CHRISTINE

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 754 433 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

Variantes de Cas para edición génica

APOYO DEL GOBIERNO

5 La presente invención se hizo con el apoyo del gobierno de EE. UU. con la subvención HR0011-11-2-0003 concedida por la Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA), subvención GM095501 concedida por los Institutos Nacionales de Salud (NIH), y subvención N66001-12-C-4207 concedida por el Centro de Sistemas de Guerra Naval y Espacial (SPAWAR). El gobierno tiene ciertos derechos en la presente invención.

ANTECEDENTES DE LA INVENCION

10 La edición dirigida de secuencias de ácidos nucleicos, por ejemplo, la introducción de una modificación específica en ADN genómico, es un enfoque muy prometedor para el estudio de la función génica y también tiene el potencial de proporcionar nuevas terapias para enfermedades genéticas humanas.¹ Una tecnología ideal de edición de ácidos nucleicos posee tres características: (1) alta eficiencia de instalación de la modificación deseada; (2) actividad inespecífica mínima; y (3) la capacidad de ser programada para editar con precisión cualquier sitio en un ácido nucleico dado, por ejemplo, cualquier sitio dentro del genoma humano.² Las actuales herramientas de ingeniería del genoma, que incluyen nucleasas de dedos de cinc modificadas (ZFNs),³ nucleasas efectoras de tipo activador de la transcripción (TALENs),⁴ y lo más recientemente, la ADN endonucleasa guiada por ARN Cas9⁵, efectúan la escisión de ADN específica de secuencia en un genoma. Esta escisión programable puede dar como resultado la mutación del ADN en el sitio de escisión mediante unión de extremos no homólogos (NHEJ) o sustitución del ADN que rodea el sitio de escisión mediante reparación dirigida por homología (HDR).^{6,7}

20 Un inconveniente de las tecnologías actuales es que tanto NHEJ como HDR son procesos estocásticos que normalmente dan como resultado modestas eficiencias de la edición génica, así como alteraciones génicas no deseadas, que pueden competir con la alteración deseada.⁸ Puesto que muchas enfermedades genéticas, en principio, se pueden tratar efectuando un cambio de nucleótidos específicos en una localización específica en el genoma (por ejemplo, un cambio de C a T en un codón específico de un gen asociado a una enfermedad),⁹ el desarrollo de una forma programable para lograr dicha edición génica con precisión representaría tanto una herramienta de investigación nueva poderosa, además de un posible nuevo enfoque de los agentes terapéuticos humanos basados en edición génica.

SUMARIO DE LA INVENCION

30 El sistema de repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) es un sistema inmunitario adaptativo procarionota¹⁰ recientemente descubierto que ha sido modificado para permitir la ingeniería del genoma robusta y general en una variedad de organismos y líneas celulares.¹¹ Los sistemas CRISPR-Cas (asociados a CRISPR) son complejos de proteína-ARN que usan una molécula de ARN (ARNgu) como guía para confinar el complejo a una secuencia de ADN diana mediante apareamiento de bases.¹² En los sistemas naturales, una proteína Cas actúa entonces de endonucleasa para escindir la secuencia de ADN dirigida.¹³ La secuencia de ADN dirigida debe ser tanto complementaria al ARNgu, como también contener un dinucleótido de "motivo adyacente de protoespaciador" (PAM) en el extremo 3' de la región complementaria con el fin de que el sistema funcione (Figura 1).¹⁴ Entre las proteínas Cas conocidas, Cas9 de *S. pyogenes* ha sido generalmente ampliamente usada como una herramienta para ingeniería del genoma.¹⁵ Esta proteína Cas9 es una proteína multi-dominio grande que contiene dos dominios de nucleasa distintos. Se pueden introducir mutaciones puntuales en Cas9 para suprimir la actividad de nucleasa, dando como resultado una Cas9 muerta (dCas9) que todavía retiene su capacidad de unirse a ADN en un modo programado por ARNgu.¹⁶ En principio, cuando se fusiona con otra proteína o dominio, dCas9 puede dirigir esa proteína a prácticamente cualquier secuencia de ADN simplemente por co-expresión con un ARNgu apropiado.

45 El potencial del complejo de dCas9 para fines de ingeniería del genoma es inmenso. Su capacidad única para llevar proteínas a sitios específicos en un genoma programado por el ARNgu en teoría se puede desarrollar en una variedad de herramientas de ingeniería del genoma específicas de sitio más allá de las nucleasas, que incluyen activadores transcripcionales, represores transcripcionales, proteínas modificadoras de histonas, integrasas y recombinasas.¹¹ Algunas de estas posibles aplicaciones se han implementado recientemente mediante fusiones de dCas9 con activadores transcripcionales para proporcionar activadores transcripcionales guiados por ARN,^{17,18} represores transcripcionales,^{16,19,20} y enzimas de modificación de cromatina.²¹ La simple co-expresión de estas fusiones con una variedad de ARNgu da como resultado la expresión específica de los genes diana. Estos estudios seminales han sentado las bases para el diseño y la construcción de efectores específicos de secuencia fácilmente programables para la precisa manipulación de genomas.

55 Significativamente, el 80-90 % de las mutaciones de proteína responsables de la enfermedad humana surgen de la sustitución, delección o inserción de solo un único nucleótido.⁶ Sin embargo, hasta ahora no se ha desarrollado ninguna herramienta de ingeniería del genoma que permita la manipulación de un único nucleótido de una manera general y directa. Las actuales estrategias para la corrección de genes de bases individuales incluyen nucleasas manipuladas (que se basan en la creación de roturas de cadena doble, DSBs, seguido por reparación dirigida por

homología, HDR, ineficiente estocástica) y oligonucleótidos quiméricos de ADN-ARN.²² La última estrategia implica el diseño de una secuencia de ARN/ADN para aparear bases con una secuencia específica en ADN genómico, excepto en el nucleótido a editar. El desapareamiento resultante es reconocido por el sistema de reparación endógeno de células y reparado, conduciendo a un cambio en la secuencia de cualquiera de la quimera o el genoma. Ambas de estas estrategias padecen bajas eficiencias de edición génica y alteraciones génicas no deseadas, ya que se someten a tanto a la estocasticidad de HDR como a la competición entre HDR y unión de extremos no homólogos, NHEJ.²³⁻²⁵ Las eficiencias de HDR varían según la localización del gen diana dentro del genoma,²⁶ el estado del ciclo celular,²⁷ y el tipo de célula/tejido.²⁸ El desarrollo de una forma programable directa para instalar un tipo específico de modificación de base en una localización precisa en el ADN genómico con eficiencia de tipo enzima y sin estocasticidad representaría, por tanto, un enfoque nuevo y poderoso a las herramientas de investigación basadas en edición génica y agentes terapéuticos humanos.

Algunos aspectos de la presente divulgación proporcionan estrategias, sistemas, reactivos, métodos y kits que son útiles para la edición de ácidos nucleicos dirigidos, que incluyen editar un único sitio dentro del genoma de un sujeto, por ejemplo, el genoma humano. Se proporcionan proteínas de fusión de Cas9 y enzimas de edición de ácidos nucleicos o dominios de enzima, por ejemplo, dominios de desaminasa. Se proporcionan métodos de edición de ácidos nucleicos dirigidos. Se proporcionan reactivos y kits para la generación de proteínas de edición de ácidos nucleicos dirigidos, por ejemplo, proteínas de fusión de Cas9 y enzimas o dominios que editan ácidos nucleicos.

En un primer aspecto, la presente invención proporciona una proteína de fusión que comprende (i) un dominio de Cas9 de nucleasa inactiva; y (ii) un dominio de desaminasa. En algunas realizaciones, la desaminasa es una citidina desaminasa. En algunas realizaciones, la desaminasa es una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC). En algunas realizaciones, la desaminasa es una desaminasa de la familia APOBEC 1. En algunas realizaciones, la desaminasa es una citidina desaminasa inducida por activación (AID). En algunas realizaciones, la desaminasa es una ACF1/ASE desaminasa. En algunas realizaciones, la desaminasa es una adenosina desaminasa. En algunas realizaciones, la desaminasa es una desaminasa de la familia ADAT. En algunas realizaciones, el dominio de edición de ácidos nucleicos se fusiona con el extremo N del dominio de Cas9. En algunas realizaciones, el dominio de edición de ácidos nucleicos se fusiona con el extremo C del dominio de Cas9. En algunas realizaciones, el dominio de Cas9 y el dominio de edición de ácidos nucleicos se fusionan por un conector. En algunas realizaciones, el conector comprende un motivo (GGGGS)_n (SEQ ID NO: 91), (G)_n, (EAAAK)_n (SEQ ID NO: 5), (GGS)_n, SGSETPGTSESATPES (SEQ ID NO: 93) (véase, por ejemplo, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82), o un motivo (XP)_n, o una combinación de cualquiera de estos, en donde n es independientemente un número entero entre 1 y 30.

En un segundo aspecto, la invención proporciona una proteína de fusión según el primer aspecto de la invención, para su uso en medicina.

En un tercer aspecto, la invención proporciona un método *in vitro* o *ex vivo* de edición de ADN, comprendiendo el método poner en contacto una molécula de ADN con (a) una proteína de fusión según el primer aspecto de la invención; y (b) un ARN guía único (ARNgu) que dirige la proteína de fusión de (a) a una secuencia de nucleótidos diana de la molécula de ADN; en donde la molécula de ADN se pone en contacto con la proteína de fusión y el ARNgu en una cantidad eficaz y en condiciones adecuadas para la desaminación de una base nucleotídica de la molécula de ADN. En algunas realizaciones, la secuencia de ADN diana comprende una secuencia asociada a una enfermedad o trastorno, y en donde la desaminación de la base nucleotídica da como resultado una secuencia que no se asocia a una enfermedad o trastorno. En algunas realizaciones, la secuencia de ADN comprende una mutación puntual T>C o A>G asociada a una enfermedad o trastorno, y en donde la desaminación de la base C o G mutante da como resultado una secuencia que no se asocia a una enfermedad o trastorno. En algunas realizaciones, la desaminación corrige una mutación puntual en la secuencia asociada a la enfermedad o trastorno. En algunas realizaciones, la secuencia asociada a la enfermedad o trastorno codifica una proteína, y en donde la desaminación introduce un codón de terminación en la secuencia asociada a la enfermedad o trastorno, dando como resultado una truncación de la proteína codificada. En algunas realizaciones, la desaminación corrige una mutación puntual en el gen PI3KCA, corrigiendo así una mutación H1047R y/o A3140G. En algunas realizaciones, la enfermedad o trastorno es una enfermedad asociada a una mutación puntual, o una mutación de una base individual, en el genoma. En algunas realizaciones, la enfermedad es una enfermedad genética, un cáncer, una enfermedad metabólica, o una enfermedad de almacenamiento lisosómico.

En un cuarto aspecto, la invención proporciona la proteína de fusión del primer aspecto de la invención para su uso en un método de tratamiento de una enfermedad o trastorno, comprendiendo el método poner en contacto una molécula de ADN con

(a) la proteína de fusión; y

(b) un ARNgu que dirige la proteína de fusión de (a) a una secuencia de ADN diana de la molécula de ADN; en donde la molécula de ADN se pone en contacto con la proteína de fusión y el ARNgu en una cantidad eficaz y en condiciones adecuadas para la desaminación de una base nucleotídica de la molécula de ADN,

en donde la secuencia de ADN diana comprende una secuencia asociada a una enfermedad o trastorno, la desaminación de la base nucleotídica da como resultado una secuencia que no se asocia a una enfermedad o trastorno, y en donde el contacto es *in vivo* en un sujeto susceptible a tener, que tiene, o diagnosticado con una enfermedad o trastorno. En algunas realizaciones, la secuencia de ADN comprende una mutación puntual T>C o A>G asociada a una enfermedad o trastorno, y en donde la desaminación de la base C o G mutante da como resultado una secuencia que no se asocia a una enfermedad o trastorno. En algunas realizaciones, la desaminación corrige una mutación puntual en la secuencia asociada a la enfermedad o trastorno. En algunas realizaciones, la secuencia asociada a la enfermedad o trastorno codifica una proteína, y en donde la desaminación introduce un codón de terminación en la secuencia asociada a la enfermedad o trastorno, dando como resultado una truncación de la proteína codificada. En algunas realizaciones, la desaminación corrige una mutación puntual en el gen PI3KCA, corrigiendo así una mutación H1047R y/o A3140G. En algunas realizaciones, la enfermedad o trastorno es una enfermedad asociada a una mutación puntual, o una mutación de una base individual, en el genoma. En algunas realizaciones, la enfermedad es una enfermedad genética, un cáncer, una enfermedad metabólica, o una enfermedad de almacenamiento lisosómico.

Algunos aspectos de la presente divulgación proporcionan una construcción indicadora para detectar actividad de edición de ácidos nucleicos de una proteína de fusión de Cas9: dominio de edición de ADN. La construcción puede comprender (a) un gen indicador que comprende un sitio diana para la proteína de edición de ADN de Cas9, en donde la edición dirigida de ADN da como resultado un aumento en la expresión del gen indicador; y (b) una secuencia promotora que controla la expresión del gen indicador. La construcción puede comprender además (c) una secuencia que codifica un ARN_g que dirige la proteína de edición de ADN de Cas9 al sitio diana del gen indicador, en donde la expresión del ARN_g es independiente de la expresión del gen indicador. El sitio diana del gen indicador comprende un codón de terminación prematuro, y en donde la edición dirigida de ADN de la cadena molde por la proteína de edición de ADN de Cas9 da como resultado una conversión del codón de terminación prematuro en un codón que codifica un resto de aminoácido. El gen indicador puede codificar una luciferasa, una proteína fluorescente, o un marcador de resistencia a antibióticos.

Algunos aspectos de la presente divulgación proporcionan kits que comprenden una construcción de ácidos nucleicos que comprende una secuencia que codifica una secuencia de Cas9 inactiva en nucleasa, una secuencia que comprende un sitio de clonación situado para permitir la clonación de una secuencia que codifica una enzima de edición de ácidos nucleicos o dominio de enzima en marco con la secuencia codificante de Cas9, y, opcionalmente, una secuencia que codifica un conector situado entre la secuencia codificante de Cas9 y el sitio de clonación. Además, el kit puede comprender reactivos, tampones, y/o instrucciones adecuados para la clonación en marco de una secuencia que codifica una enzima de edición de ácidos nucleicos o dominio de enzima en la construcción de ácidos nucleicos para generar una proteína de fusión de edición de ácidos nucleicos de Cas9. La secuencia que comprende el sitio de clonación puede ser el extremo N de la secuencia de Cas9. La secuencia que comprende el sitio de clonación puede ser el extremo C de la secuencia de Cas9. El conector codificado puede comprender un motivo (GGGGS)_n (SEQ ID NO: 91), (G)_n, (EAAAK)_n(SEQ ID NO: 5), (GGG)_n, SGSETPGTSESATPES (SEQ ID NO: 93) (véase, por ejemplo, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82, o un motivo (XP)_n, o una combinación de cualquiera de estos, en donde n es independientemente un número entero entre 1 y 30.

Algunos aspectos de la presente divulgación proporcionan kits que comprenden una proteína de fusión que comprende un dominio de Cas9 de nucleasa inactiva y un enzima de edición de ácidos nucleicos o dominio de enzima, y, opcionalmente, un conector situado entre el dominio de Cas9 y la enzima de edición de ácidos nucleicos o dominio de enzima. Además, el kit puede comprender reactivos, tampones y/o instrucciones adecuados para usar la proteína de fusión, por ejemplo, para edición *in vitro* o *in vivo* de ADN o ARN. El kit puede comprender instrucciones referentes al diseño y uso de ARN_g adecuadas para la edición dirigida de una secuencia de ácidos nucleicos.

El sumario anterior se indica para ilustrar, de una manera no limitante, algunas de las realizaciones, ventajas, características y usos de la tecnología desvelada en el presente documento. Otras realizaciones, ventajas, características y usos de la tecnología desvelada en el presente documento serán evidentes a partir de la descripción detallada, los dibujos, los ejemplos y las reivindicaciones.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Figura 1. El complejo de Cas9/ARN_g-ADN. El extremo 3' del ARN_g forma un complejo de ribonucleoproteína con la nucleasa Cas9, mientras que el extremo 5' de 20 nt del ARN_g reconoce su extensión complementaria de ADN. La unión de ADN requiere la secuencia 5' de PAM de 3 nt para el ADN diana. En el caso de wtCas9, ocurre la escisión de ADN de cadena doble 3 nt desde PAM para producir extremos romos (mostrados por las flechas). Se debe observar que el tamaño de la burbuja es desconocido.

Figura 2. Estructura cristalina del dominio catalítico de APOBEC3G (PDB ID 3E1U). La estructura secundaria de núcleo, que se cree que está conservada entre toda la familia, consiste en una lámina β de cinco cadenas (flechas) flanqueada por seis hélices α. Se cree que el bucle de centro activo (bucle de sitio activo) es

responsable de determinar la especificidad de desaminación. El Zn^{2+} responsable de la actividad catalítica se muestra como una esfera. Las secuencias corresponden, de arriba a abajo, a SEQ ID NOs: 97-98.

Figura 3. Diseño de ensayo indicador basado en luciferasa. Se variará el ARN_g para dirigir numerosas secuencias que corresponden a regiones antes de y que incluyen el gen luciferasa para dirigir el codón de iniciación mutado (resto C subrayado). Se añadirá una región "tampón" entre el codón de iniciación y el gen luciferasa para incluir codones de solo A y T (mostrados como (ZZZ)_x). Se indica la secuencia de Shine-Dalgarno. En algunas realizaciones, es preferible mantener todos los C apareados con bases para prevenir efectos inespecíficos.

Figura 4. Ensayo de desaminasa. Las secuencias corresponden, de arriba a abajo, a SEQ ID NOs: 99-105.

Figura 5. Gel de SDS-PAGE de ADN_{bc} editado por proteínas de fusión Cas9-APOBEC1.

DEFINICIONES

Como se usa en el presente documento y en las reivindicaciones, las formas en singular "un", "una", "el" y "la" incluyen la referencia al singular y al plural, a menos que el contexto indique claramente de otro modo. Así, por ejemplo, una referencia a "un agente" incluye un único agente y una pluralidad de dichos agentes.

El término "Cas9" o "nucleasa Cas9" se refiere a una nucleasa guiada por ARN que comprende una proteína Cas9, o un fragmento de la misma (por ejemplo, una proteína que comprende un dominio de escisión de ADN activo o inactivo de Cas9, y/o el dominio de unión de ARN_g de Cas9). Una nucleasa Cas9 también se refiere algunas veces como una nucleasa *casn1* o una nucleasa asociada a CRISPR (repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas). CRISPR es un sistema inmunitario adaptativo que proporciona protección contra elementos genéticos móviles (virus, elementos transponibles y plásmidos conjugativos). Las agrupaciones CRISPR contienen espaciadores, secuencias complementarias a elementos móviles antecedentes, y se dirigen a ácidos nucleicos invasores. Las agrupaciones de CRISPR se transcriben y procesan en ARN de CRISPR (ARN_{cr}). En sistemas CRISPR de tipo II, el correcto procesamiento de pre-ARN_{cr} requiere un ARN pequeño trans-codificado (ARN_{cr}_{tra}), ribonucleasa endógena 3 (*rnc*) y una proteína Cas9. El ARN_{cr}_{tra} sirve de guía para el procesamiento ayudado por ribonucleasa 3 de pre-ARN_{cr}. Posteriormente, Cas9/ARN_{cr}/ARN_{cr}_{tra} escinde endonucleolíticamente el ADN_{bc} diana lineal o circular complementario al espaciador. La cadena diana no complementaria a ARN_{cr} se corta primero endonucleolíticamente, luego se corta en 3'-5' exonucleolíticamente. En la naturaleza, la unión y escisión de ADN normalmente requiere proteína y ambos ARNs. Sin embargo, se pueden manipular los ARN guía único ("ARN_g", o simplemente "ARN_g") para incorporar aspectos de tanto el ARN_{cr} como el ARN_{cr}_{tra} en una única especie de ARN. Véase, por ejemplo, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012). Cas9 reconoce un motivo corto en las secuencias de repetición de CRISPR (PAM o el motivo adyacente de protoespaciador) para ayudar a distinguir propio frente a no propio. Las secuencias y estructuras de la nucleasa Cas9 son bien conocidas por los expertos en la técnica (véanse, por ejemplo, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); y "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012)). Se han descrito ortólogos de Cas9 en diversas especies, que incluyen, pero no se limitan a, *S. pyogenes* y *S. thermophilus*. Las nucleasas y secuencias Cas9 adecuadas adicionales serán evidentes para los expertos en la técnica basándose en la presente divulgación, y dichas nucleasas y secuencias de Cas9 incluyen secuencias de Cas9 de los organismos y loci desvelados en Chylinski, Rhun y Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737. En algunas realizaciones, una nucleasa Cas9 tiene un dominio de escisión de ADN inactivo (por ejemplo, uno inactivado).

Una proteína Cas9 inactivada en nucleasa se puede denominar indistintamente una proteína "dCas9" (de nucleasa Cas9 "muerta"). Se conocen métodos de generación de una proteína Cas9 (o un fragmento de la misma) que tienen un dominio de escisión de ADN inactivo (véase, por ejemplo, Jinek et al., *Science*. 337:816-821(2012); Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28;152(5):1173-83. Por ejemplo, se conoce que el dominio de escisión de ADN de Cas9 incluye dos subdominios, el dominio de nucleasa HNH y el subdominio RuvC1. El subdominio HNH escinde la cadena complementaria al ARN_g, mientras que el subdominio RuvC1 escinde la cadena no complementaria. Las mutaciones dentro de estos subdominios pueden silenciar la actividad de nucleasa de Cas9. Por ejemplo, las mutaciones D10A y H841A inactivan completamente la actividad de nucleasa de Cas9 de *S. pyogenes* (Jinek et al., *Science*. 337:816-821(2012); Qi et al., *Cell*. 28;152(5):1173-83 (2013). En algunas realizaciones, se proporcionan proteínas que comprenden fragmentos de Cas9. Por ejemplo, en algunas realizaciones, una proteína comprende uno de los dos dominios de Cas9: (1) el dominio de unión ARN_g de Cas9; o (2) el dominio de escisión de ADN de Cas9. En algunas realizaciones, las proteínas que comprenden Cas9 o fragmentos de la misma se denominan "variantes de Cas9". Una variante de Cas9 comparte homología con Cas9, o un fragmento de la misma. Por ejemplo, una variante de

5 Cas9 es al menos aproximadamente 70 % idéntica, al menos aproximadamente 80 % idéntica, al menos aproximadamente 90 % idéntica, al menos aproximadamente 95 % idéntica, al menos aproximadamente 96 % idéntica, al menos aproximadamente 97 % idéntica, al menos aproximadamente 98 % idéntica, al menos aproximadamente 99 % idéntica, al menos aproximadamente 99,5 % idéntica, o al menos aproximadamente 99,9 % a Cas9 no mutante. En algunas realizaciones, la variante de Cas9 comprende un fragmento de Cas9 (por ejemplo, un dominio de unión de ARNg o un dominio de escisión de ADN), de forma que el fragmento sea al menos aproximadamente 70 % idéntico, al menos aproximadamente 80 % idéntico, al menos aproximadamente 90 % idéntico, al menos aproximadamente 95 % idéntico, al menos aproximadamente 96 % idéntico, al menos aproximadamente 97 % idéntico, al menos aproximadamente 98 % idéntico, al menos aproximadamente 99 % idéntico, al menos aproximadamente 99,5 % idéntico, o al menos aproximadamente 99,9 % al fragmento correspondiente de Cas9 no mutante. En algunas realizaciones, Cas9 no mutante corresponde a Cas9 de *Streptococcus pyogenes* (secuencia de referencia de NCBI:

NC_017053.1, SEQ ID NO: 1 (nucleótido); SEQ ID NO: 2 (aminoácido)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCCGGATGGGCGGTGATCACTGATGATTAT
AAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCCACAGTATCAAAAAAATCTTATAGGGGCT
CTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG
AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCACTCTATTTTGGAAAATATAGTAGATGAA
GTTGCTTATCATGAGAAAATATCCAATATCTATCATCTGCGAAAAAATTTGGCAGATTCTACTGATAAAGCGGAT
TTGGCGTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCAATTTTTGATTGAGGGAGATTTAAAT
CCTGATAAATAGTGATGTGGACAAACTATTTTATCCAGTTGGTACAAAATCTACAATCAATTTATTTGAAGAAAACCT
ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC
ATTGCTCAGCTCCCCGGTGAAGAAGAAAATGGCTTGTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCT
AATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATACAGCTTTCAAAAGATACTTACGATGATGATTTA
GATAATTTATTTGGCGCAAATTTGGAGATCAATATGCTGATTTGTTTTGGCAGCTAAGAATTTATCAGATGCTATT
TTACTTTACAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAAGCGCTAC
GATGAACAATCAAGACTTACTCTTTTTAAAGCTTTAGTTTCGACAACTTCCAGAAAAGTATAAAGAAAATC
TTTTTTGATCAATCAAAAAACGGATATGCAGTTATATTGATGGGGGAGCTAGCCAAGAAGAAATTTTATAAATTT
ATCAAACCAATTTTAGAAAAATGGATGGTACTGAGGAATTTATGGTGAAGTAAATCGTGAAGATTGCTGCGC
AAGCAACGGACCTTTGACAACGGCTCTATTTCCCATCAAATTCCTTTGGGTGAGCTGCATGCTATTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTAT
TATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGTCATGGATGACTCGGAAGTCTGAAGAAACAATTAACCCCA
TGGAATTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTTGAACGCATGACAAAATTTGATAAA
AATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGTCTTTATGATATTTTACGGTTTATAACGAATTGACA
AAGGTCAAATATGTTACTGAGGGAAATGCGAAAACAGCATTTCTTTTCAAGTGAACAGAAGAAAGCCATTGTTGAT
TTACTCTTCAAACAATCGAAAAGTAACCGTTAAGCAATTTAAAGAAGATTATTTCAAAAAATAGAATGTTTTT
GATAGTGTGAAAATTTCAAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCACCATTGATTTGCTAAAAAT
ATTAAGAATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTA
TTTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATGATTAATGGTATTAGGGATAAGCAATCT
GGCAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT
AGTTTGACATTTAAAGAAGATATTTCAAAAAGCACAGGTGCTGGACAAGGCCATAGTTTACATGAACAGATTGCT
AACTTAGCTGGCAGTCTGCTATTTAAAAAGGTATTTTTACAGACTGTAAAAATTTGTTGATGAACTGGTCAAAGTA
ATGGGGCAATAAGCCAGAAAATATCGTTATTGAAATGGCAGCTGAAAATCAGACAACCTCAAAGGGCCAGAAAAAT
TCGCGAGAGCGTATGAAACGAATCGAAGAAGTATCAAAGAATTAGGAAGTCAGATTTCTTAAAGAGCATCCTGTT
GAAAATACCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAA
GAATTAGATATTAATCGTTTAAAGTATTATGATGTCGATCAGATTTGTTCCACAAAGTTTCATTTAAAGACGATTCA
ATAGACAAATAAGGTACTAACCGGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC
AAAAAGATGAAAACATTTGGAGACAACCTTCAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACG
AAAGCTGAACGTTGGAGGTTTGGAGTGAACCTTGAATAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA
ATCACTAAGCATGTGGCACAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTTCGA
GAGGTTAAAGTATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTCCAATTTCTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTGGAAGTCTTTGATTAAGAAAATAT
CCAAAATTTGAATCGGAGTTTGTCTATGGTGAATATAAAGTTTATGATGTTGCTAAAATGATTGCTAAGTCTGAG
CAAGAAATAGGCAAAGCAACCGCAAATATTTCTTTTACTCTAATATCATGAACCTTTTCAAACAGAAAATTTACA
CTTGCAAAATGGAGAGATTCGCAAACGCCCTCTAATCGAACTAATGGGAAAACCTGGAGAAAATTTGCTGGGATAAA
GGCGGAGATTTGCCACAGTGGCAAAAGTATTGTCCATGCCCAAGTCAATTTGTCAGAAAACAGAAGTACAG
ACAGGCGGATTTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATGCTCGTAAAAAAGACTGG

ES 2 754 433 T3

GATCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCTAGTGGTTGCTAAGGTGGAAAAA
GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAAATATGGAAAAGAAGTTCCTTTGAAAAA
AATCCGATTGACTTTTTAGAAAGCTAAAGGATATAAGGAAGTTAAAAAGACTTAATCATTAAACTACCTAAATAT
AGTCTTTTTGAGTTAGAAAACGGTCGTAACGGATGCTGGCTAGTGCCGAGAAATTACAAAAAGGAAATGAGCTG
GCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAAGAT
AACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAATCAGTGAATTT
TCTAAGCGTGTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATAT
TTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTATAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:1)

MDKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFGSGETAETRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLADSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRENLE
LAQLPGEKRNGLFGNLLALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLRLKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGRMKAFLSSEGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLKI
IKDKDFLDNEENEDILEDIVLTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQS
GKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGOGHSLHEQIANLAGSPAIKKGIILQTVKIVDELVKV
MGHKPENIVIEMARENQTTQKGQNSRERMKRIEEGIKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFIKDSDIDNKVLRSDKNRKGSDNVPSEEVVKKMKNYWRQLLNKLIITQRKFDNLT
KAERGGLELDKAGFTKRQLVETRQITKHYAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKIVDVRKMIKSEOEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKYYGGFDSPTVAYSVLVVAKVEKGSKLLKSVKELLGTTIMERSSEFEKNPIDFLEAKGYKEVKKDLI IKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEIQISEF
SKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
ITGLYETRIDLSQLGGD (SEQ ID NO:2)

(subrayado sencillo: dominio HNH; subrayado doble: dominio RuvC)

En algunas realizaciones, Cas9 no mutante corresponde a, o comprende, SEQ ID NO: 3 (nucleótido) y/o SEQ ID NO: 4 (aminoácido):

5

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACATAATCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGTGTGGGGAACACAGACCGTCAATTCGATTAAAAGAATCTATCGGTGCC
CTCCTATTTCGATATGGCGAAACGGCAGAGGGCAGCTCGCCTGAAACGAACCGCTCGGAGAAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGTAGGGCAAAGTTGACGATTTCTTTTACCGGT
TTGGAAGAGTCTCTCTTGTGCGAAGAGGACAAGAAACATGAACGGCACCCCTCTTTGGAAACATATGATGAG
GTGGCATA TCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTCCGTTGGGCACTTTCTCATTGAGGGTGTATTAAT
CCGGACAACCTCGGATGTCGACAACTGTTTCAATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATCTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAAATACCCGAGAGAAGAAAAATGGGTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCCCTGACACCA
AATTTTAAGTCGAACCTCGACTTAGCTGAAGATGCCAAATTCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACGAGAGTTAATACTGAGATTACCAAGCGCCGTTATCCGCTTCAATGATCAAAAAGGTAC
GATGAACATCACCAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATCTACAAGTTT
ATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAACCACTCAATCGCGAAGATCTACTCGGA
AAGCAGCGGACTTTGACAAACGGTAGCATTCCACATCAATCCACTTAGGCGAATTCATGCTATACTTAGAAGG
CAGGAGGATTTTTATCCGTTCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTGCGATACCTTAC
TATGTGGGACCCCTGGCCGAGGGAACCTCGGTTGCGATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGTCAATCGTTTATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTTACAGTGTACAATGAACCTCACG
AAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCTTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAGACCAACCGCAAAGTGACAGTTAAGCAATTTGAAAGAGGACTACTTTAAGAAAATGAATGCTTC
GATTCGTGCGAGATCTCCGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAGATAAGGACTTCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCCTC

ES 2 754 433 T3

TTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAG
TTAAAGAGCGCTCGCTATACGGGCTGGGGACGATTGTTCGGGAAACTTATCAACGGGATAAGAGACAAGCAAAGT
GGTAAAAC TATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGC TGGTTCGCCAGCCATCAAAAAGGCATACCTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCAGCGGAAAATCAAACGACTCAGAAGGGGCAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAGAAGTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATGCAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAATCGACATAAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAAGTGACAAATGTTCCAGCGGAGGAAATC
GTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAAATGATAACGCAAAGAAAAGTTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAAC TTGACAAGCCGGATTTATTAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAAATACGACGAGAAAGCATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTCCGACTTCAGAAAAGGATTTTCAATTTCTATAAAGTT
AGGGAGATAAATAACTACCACATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCCGACTCATTAAAGAAA
TACCCGAAGCTAGAAAAGTGAAGTTTGTGTATGGTGATTACAAAGTTTATGACGCTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAAGTGGAGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTCCAAAAGGAATAGTGA TAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATCTGTCTTAGTAGTGGCAAAAAGTTGAG
AAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCATCGACTTCCTTGAGGCGAAAGTTACAAGGAAGTAAAAAGGATCTCATAATTAACCTACCAAAG
TATAGTCTGTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAGGGGAAACGAA
CTCGCACTACCGTCTAAAATACGTGAATTTCTGTATTTAGCGTCCCATACGAGAAGTTGAAAGGTTCACTGAA
GATAACGAACAGAAAGCAACTTTTTGTTGAGCAGCACAAAACATTATCTCGACGAAATCATAGAGCAAATTTTCGGAA
TTCAGTAAGAGAGTCACTCTAGCTGATGCCAATCTGGACAAAAGTATTAAGCGCATAACAACAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAA
TCCATCACGGGATTATATGAAACTCGGATAGATTTGTACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGG
AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGCAGGA (SEQ ID NO:3)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHS IKKNLIGALLEDSGETAEATRLKRTARRRYTRR
RNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLVDS TKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENL
IAQLPGEKKNLFGNLLALSGLTPNFKSNFLAEDAKLQLSKDTYDDLDLNLQAIGDQYADLFLAANKLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDL LRKQRTFDNGS IPHQIHLGELHAI LRRQEDFY PFLKDNREKIEKILTRIPY
YVGPLARGNSRFWMTRKSEETITPWNFEVVDKGASQSF IERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELT
KVYVTEGRMKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDI LEDI VLTLTFEDREMIERLKYAHLFDKVMKQLKRRRYTGWRLSRKLINGIRDKQS
GKTI LDFLKS DGFANRNFQLIHDDSLTFKEDI QKAQVSGQDSLHEHIANLAGSPA I KKG I LQTVKVVDELVKV
MGRHKPENIVLEMARENQTTQKQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRGS DNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFKRQLVETROITKHVAQILDSRMTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMI AKSEQEIGKATAKYFFYSNIMNFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIYKKTEVQTGGF SKESILPKRNSDKLIARKKD
WDPKKGFFDSPTVAVSVLVVAKVEKGSKKLKS VKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLI IKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVLADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:4)

(subrayado sencillo: dominio HNH; subrayado doble: dominio RuvC)

5 En algunas realizaciones, dCas9 corresponde a, o comprende en parte o en total, una secuencia de aminoácidos de Cas9 que tiene una o más mutaciones que inactivan la actividad de la nucleasa Cas9. Por ejemplo, en algunas realizaciones, un dominio de dCas9 comprende la mutación D10A y/o H820A.

dCas9 (D10A y H840A):

MDKKYSIGLAIGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKHKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TKADLRLIYLALAHMIKFRGHFLIEGDLNPNDSVDVKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHOIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPAIKKGILOTVKVVDELVKV
MGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVAIVPQSFLKDDSIDNKVLRSDKNRGSNDVPSSEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKAKSEQIEGKATAKYFFYSNIMNFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGSKLLKSVKELLGITIMERSSEKPNIDFLEAKGYEVKDLI IKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 34)

(subrayado sencillo: dominio HNH; subrayado doble: dominio RuvC)

5 En otras realizaciones, se proporcionan variantes de dCas9 que tienen mutaciones distintas de D10A y H820A, que, por ejemplo, dan como resultado Cas9 inactivada en nucleasa (dCas9). Dichas mutaciones, a modo de ejemplo, incluyen otras sustituciones de aminoácidos en D10 y H820, u otras sustituciones dentro de los dominios de nucleasa de Cas9 (por ejemplo, sustituciones en el subdominio de nucleasa HNH y/o el subdominio RuvC1). En algunas realizaciones, se proporcionan variantes u homólogos de dCas9 (por ejemplo, variantes de SEQ ID NO: 34) que son al menos aproximadamente 70 % idénticas, al menos aproximadamente 80 % idénticas, al menos aproximadamente 90 % idénticas, al menos aproximadamente 95 % idénticas, al menos aproximadamente 98 % idénticas, al menos aproximadamente 99 % idénticas, al menos aproximadamente 99,5 % idénticas, o al menos aproximadamente 99,9 % a SEQ ID NO: 34. En algunas realizaciones, se proporcionan variantes de dCas9 (por ejemplo, variantes de SEQ ID NO: 34) que tienen secuencias de aminoácidos que son más cortas, o más largas que SEQ ID NO: 34, por aproximadamente 5 aminoácidos, por aproximadamente 10 aminoácidos, por aproximadamente 15 aminoácidos, por aproximadamente 20 aminoácidos, por aproximadamente 25 aminoácidos, por aproximadamente 30 aminoácidos, por aproximadamente 40 aminoácidos, por aproximadamente 50 aminoácidos, por aproximadamente 75 aminoácidos, por aproximadamente 100 aminoácidos o más.

20 En algunas realizaciones, las proteínas de fusión de Cas9 como se proporcionan en el presente documento comprenden el aminoácido de longitud completa de una proteína Cas9, por ejemplo, una de las secuencias proporcionadas anteriormente. En otras realizaciones, sin embargo, las proteínas de fusión como se proporcionan en el presente documento no comprenden una secuencia de Cas9 de longitud completa, sino solo un fragmento de la misma. Por ejemplo, en algunas realizaciones, una proteína de fusión de Cas9 proporcionada en el presente documento comprende un fragmento de Cas9, en donde el fragmento se une a ARNcr y ARNcrtra o ARNcu, pero no comprende un dominio de nucleasa funcional, por ejemplo, en el que comprende solo una versión truncada de un dominio de nucleasa o dominio de no nucleasa en absoluto. Se proporcionan en el presente documento secuencias de aminoácidos a modo de ejemplo de dominios de Cas9 y fragmentos de Cas9 adecuados, y dominios y fragmentos de secuencias de Cas9 adecuados adicionales serán evidentes para los expertos en la técnica.

30 En algunas realizaciones, Cas9 se refiere a Cas9 de: *Corynebacterium ulcerans* (Ref. de NCBI: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (Ref. de NCBI: NC_016782.1, NC_016786.1); *Spiroplasma syrrhodicola* (Ref. de NCBI: NC_021284.1); *Prevotella intermedia* (Ref. de NCBI: NC_017861.1); *Spiroplasma taiwanense* (Ref. de NCBI: NC_021846.1); *Streptococcus iniae* (Ref. de NCBI: NC_021314.1); *Belliella baltica* (Ref. de NCBI: NC_018010.1); *Psychroflexus torquis* (Ref. de NCBI: NC_018721.1); *Streptococcus thermophilus* (Ref. de NCBI: YP_820832.1); *Listeria innocua* (Ref. de NCBI: NP_472073.1); *Campylobacter jejuni* (Ref. de NCBI: YP_002344900.1); o *Neisseria meningitidis* (Ref. de NCBI: YP_002342100.1).

35 El término "desaminasa" se refiere a una enzima que cataliza una reacción de desaminación. En algunas realizaciones, la desaminasa es una citidina desaminasa, que cataliza la desaminación hidrolítica de citidina o desoxicitidina a uracilo o desoxiuracilo, respectivamente.

40 El término "cantidad eficaz", como se usa en el presente documento, se refiere a una cantidad de un agente biológicamente activo que es suficiente para provocar una respuesta biológica deseada. Por ejemplo, en algunas realizaciones, una cantidad eficaz de una nucleasa se puede referir a la cantidad de la nucleasa que es suficiente para inducir la escisión de un sitio diana específicamente unido y escindido por la nucleasa. En algunas realizaciones, una cantidad eficaz de una proteína de fusión proporcionada en el presente documento, por ejemplo, de una proteína de fusión que comprende un dominio de Cas9 de nucleasa inactiva y un dominio de edición de ácidos nucleicos (por ejemplo, un dominio de desaminasa) se puede referir a la cantidad de la proteína de fusión que es suficiente para inducir la edición de un sitio diana específicamente unido y editado por la proteína de fusión.

Como será apreciado por el experto, la cantidad eficaz de un agente, por ejemplo, una proteína de fusión, una nucleasa, una desaminasa, una recombinasa, una proteína híbrida, un dímero de proteína, un complejo de una proteína (o dímero de proteína) y un polinucleótido, o un polinucleótido, puede variar dependiendo de diversos factores como, por ejemplo, de la respuesta biológica deseada, por ejemplo, del alelo específico, genoma, o sitio diana a editar, en la célula o tejido que se dirige, y del agente que se usa.

El término "conector", como se usa en el presente documento, se refiere a un grupo químico o una molécula que une dos moléculas o restos, por ejemplo, dos dominios de una proteína de fusión, tales como, por ejemplo, un dominio de Cas9 de nucleasa inactiva y un dominio de edición de ácidos nucleicos (por ejemplo, un dominio de desaminasa). En algunas realizaciones, un conector une un dominio de unión de ARNg de una nucleasa programable por ARN, que incluye un dominio de nucleasa de Cas9, y el dominio catalítico de proteína de edición de ácidos nucleicos. En algunas realizaciones, un conector une una dCas9 y una proteína de edición de ácidos nucleicos. Normalmente, el conector está situado entre, o flanqueado por, dos grupos, moléculas, u otros restos y conectados entre sí por un enlace covalente, que conecta así los dos. En algunas realizaciones, el conector es un aminoácido o una pluralidad de aminoácidos (por ejemplo, un péptido o proteína). En algunas realizaciones, el conector es una molécula orgánica, grupo, polímero, o resto químico. En algunas realizaciones, el conector tiene 5-100 aminoácidos de longitud, por ejemplo, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 o 150-200 aminoácidos de longitud. También se contemplan conectores más largos o más cortos.

El término "mutación", como se usa en el presente documento, se refiere a una sustitución de un resto dentro de una secuencia, por ejemplo, una secuencia de ácidos nucleicos o de aminoácidos, con otro resto, o una delección o inserción de uno o más restos dentro de una secuencia. Las mutaciones se describen normalmente en el presente documento identificando el resto original seguido por la posición del resto dentro de la secuencia y por la identidad del resto recién sustituido. Se conocen bien en la técnica diversos métodos de preparación de las sustituciones de aminoácidos (mutaciones) proporcionadas en el presente documento, y se proporcionan por, por ejemplo, Green y Sambrook, *Molecular Cloning: A Laboratory Manual* (4^a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

Los términos "ácido nucleico" y "molécula de ácido nucleico", como se usa en el presente documento, se refieren a un compuesto que comprende una nucleobase y un resto ácido, por ejemplo, un nucleósido, un nucleótido, o un polímero de nucleótidos. Normalmente, los ácidos nucleicos poliméricos, por ejemplo, las moléculas de ácidos nucleicos que comprenden tres o más nucleótidos, son moléculas lineales, en las que nucleótidos adyacentes se unen entre sí por un enlace fosfodiéster. En algunas realizaciones, "ácido nucleico" se refiere a restos individuales de ácidos nucleicos (por ejemplo, nucleótidos y/o nucleósidos). En algunas realizaciones, "ácido nucleico" se refiere a una cadena de oligonucleótidos que comprende tres o más restos de nucleótidos individuales. Como se usa en el presente documento, los términos "oligonucleótido" y "polinucleótido" se pueden usar indistintamente para referirse a un polímero de nucleótidos (por ejemplo, una cadena de al menos tres nucleótidos). En algunas realizaciones, "ácido nucleico" engloba ARN, así como ADN de cadena sencilla y/o doble. Los ácidos nucleicos pueden existir de forma natural, por ejemplo, en el contexto de un genoma, un transcrito, un ARNm, ARNt, ARNr, ARNip, ARNnp, un plásmido, cósmido, cromosoma, cromátida, u otra molécula de ácido nucleico que existe de forma natural. Por otra parte, una molécula de ácido nucleico puede ser una molécula que no existe de forma natural, por ejemplo, un ADN o ARN recombinante, un cromosoma artificial, un genoma manipulado, o fragmento de los mismos, o un ADN sintético, ARN, híbrido ADN/ARN, o que incluye nucleótidos o nucleósidos que no existen de forma natural. Además, los términos "ácido nucleico", "ADN", "ARN" y/o términos similares incluyen análogos de ácido nucleico, por ejemplo, análogos que tienen aparte de esqueleto fosfodiéster. Los ácidos nucleicos se pueden purificar de fuentes naturales, producir usando sistemas de expresión recombinante y purificar opcionalmente, sintetizar químicamente, etc. Cuando corresponda, por ejemplo, en el caso de moléculas químicamente sintetizadas, los ácidos nucleicos pueden comprender análogos de nucleósido tales como análogos que tienen bases o azúcares químicamente modificados, y modificaciones de esqueleto. Una secuencia de ácidos nucleicos se presenta en la dirección 5' a 3', a menos que se indique lo contrario. En algunas realizaciones, un ácido nucleico es o comprende nucleósidos naturales (por ejemplo, adenosina, timidina, guanosa, citidina, uridina, desoxiadenosina, desoxitimidina, desoxiguanosina y desoxicitidina); análogos de nucleósidos (por ejemplo, 2-aminoadenosina, 2-tiotimidina, inosina, pirrolopirimidina, 3-metiladenosina, 5-metilcitidina, 2-aminoadenosina, C5-bromouridina, C5-fluorouridina, C5-yodouridina, C5-propinil-uridina, C5-propinil-citidina, C5-metilcitidina, 2-aminoadenosina, 7-deazaadenosina, 7-deazaguanosina, 8-oxoadenosina, 8-oxoguanosina, O(6)-metilguanina y 2-tiocitidina); bases químicamente modificadas; bases biológicamente modificadas (por ejemplo, bases metiladas); bases intercaladas; azúcares modificados (por ejemplo, 2'-fluororibosa, ribosa, 2'-desoxiribosa, arabinosa y hexosa); y/o grupos fosfato modificados (por ejemplo, fosforotioatos y enlaces 5'-N-fosforamidato).

El término "enfermedad proliferativa", como se usa en el presente documento, se refiere a cualquier enfermedad en la que la homeostasis de células o tejidos se perturba en que una célula o población de células presenta una velocidad de proliferación anormalmente elevada. Las enfermedades proliferativas incluyen enfermedades hiperproliferativas, tales como afecciones hiperplásicas pre-neoplásicas y enfermedades neoplásicas. Las enfermedades neoplásicas se caracterizan por una proliferación anormal de células e incluyen tanto neoplasias benignas como malignas. La neoplasia maligna también se denomina cáncer.

Los términos "proteína", "péptido" y "polipéptido" se usan indistintamente en el presente documento, y se refieren a un polímero de restos de aminoácidos unidos juntos por enlaces peptídicos (amida). Los términos se refieren a una proteína, péptido, o polipéptido de cualquier tamaño, estructura, o función. Normalmente, una proteína, péptido o polipéptido tendrá al menos tres aminoácidos de longitud. Una proteína, péptido o polipéptido se puede referir a una proteína individual o un conjunto de proteínas. Se pueden modificar uno o más de los aminoácidos en una proteína, péptido o polipéptido, por ejemplo, mediante la adición de una entidad química tal como un grupo hidrato de carbono, un grupo hidroxilo, un grupo fosfato, un grupo farnesilo, un grupo isofarnesilo, un grupo ácido graso, un conector para conjugación, funcionalización, u otra modificación, etc. Una proteína, péptido o polipéptido también puede ser una molécula individual o pueden ser un complejo multi-molecular. Una proteína, péptido o polipéptido puede ser solamente un fragmento de una proteína o péptido que existe de forma natural. Una proteína, péptido o polipéptido puede existir de forma natural, ser recombinante, o sintético, o cualquier combinación de los mismos. El término "proteína de fusión", como se usa en el presente documento, se refiere a un polipéptido híbrido que comprende dominios de proteína de al menos dos proteínas diferentes. Una proteína se puede localizar en la porción de extremo amino (extremo N) de la proteína de fusión o en la proteína de extremo carboxi (extremo C) formando así una "proteína de fusión de extremo amino" o una "proteína de fusión de extremo carboxi", respectivamente. Una proteína puede comprender diferentes dominios, por ejemplo, un dominio de unión de ácido nucleico (por ejemplo, el dominio de unión de ARNg de Cas9 que dirige la unión de la proteína a un sitio diana) y un dominio de escisión de ácido nucleico o un dominio catalítico de una proteína de edición de ácidos nucleicos. En algunas realizaciones, una proteína comprende una parte proteínica, por ejemplo, una secuencia de aminoácidos que constituye un dominio de unión de ácido nucleico, y un compuesto orgánico, por ejemplo, un compuesto que puede actuar de agente de escisión de ácido nucleico. En algunas realizaciones, una proteína está en un complejo con, o está en asociación con, un ácido nucleico, por ejemplo, ARN. Cualquiera de las proteínas proporcionadas en el presente documento se puede producir por cualquier método conocido en la técnica. Por ejemplo, las proteínas proporcionadas en el presente documento se pueden producir por expresión y purificación de proteínas recombinantes, que es especialmente apta para proteínas de fusión que comprenden un conector peptídico. Se conocen bien los métodos de expresión y purificación de proteínas recombinantes, e incluyen los descritos por Green y Sambrook, *Molecular Cloning: A Laboratory Manual* (4^a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

El término "nucleasa programable por ARN" y "nucleasa guiada por ARN" se usan indistintamente en el presente documento y se refieren a una nucleasa que forma un complejo con (por ejemplo, se une o se asocia con) uno o más ARN que no son una diana para la escisión. En algunas realizaciones, una nucleasa programable por ARN, cuando está en un complejo con un ARN, se puede denominar un complejo de nucleasa:ARN. Normalmente, el (los) ARN(s) unido(s) se denominan un ARN guía (ARNg). Los ARNg pueden existir como un complejo de dos o más ARNs, o como una molécula de ARN individual. Los ARNg que existen como una molécula de ARN única se pueden denominar ARN guía único (ARNgu), aunque "ARNg" se usa indistintamente para referirse a ARNs guía que existen como o moléculas individuales o como un complejo de dos o más moléculas. Normalmente, los ARNg que existen como especies de ARN únicas comprenden dos dominios: (1) un dominio que comparte homología con un ácido nucleico diana (por ejemplo, y dirige la unión de un complejo de Cas9 a la diana); y (2) un dominio que se une a la proteína Cas9. En algunas realizaciones, el dominio (2) corresponde a una secuencia conocida como una ARNcrtra, y comprende una estructura de tallo-bucle. Por ejemplo, en algunas realizaciones, el dominio (2) es homólogo a un ARNcrtra como se representa en la Figura 1E de Jinek et al., *Science* 337:816-821(2012). Otros ejemplos de ARNg (por ejemplo, los que incluyen el dominio 2) se pueden encontrar en la solicitud de patente provisional de EE. UU. U.S.S.N. 61/874.682, presentada el 6 de septiembre de 2013, titulada "Switchable Cas9 Nucleases And Uses Thereof" y la solicitud de patente provisional de EE. UU. U.S.S.N. 61/874.746, presentada el 6 de septiembre de 2013, titulada "Delivery System For Functional Nucleases". En algunas realizaciones, un ARNg comprende dos o más de los dominios (1) y (2), y se puede denominar un "ARNg extendido". Por ejemplo, un ARNg extendido se unirá, por ejemplo, a dos o más proteínas Cas9 y se unirá a ácido nucleico diana en dos o más regiones distintas, como se describe en el presente documento. El ARNg comprende una secuencia de nucleótidos que complementa un sitio diana, que media en la unión del complejo de nucleasa/ARN a dicho sitio diana, proporcionando la especificidad de secuencia del complejo de nucleasa:ARN. En algunas realizaciones, la nucleasa programable por ARN es la endonucleasa Cas9 (sistema asociado a CRISPR), por ejemplo Cas9 (Csn1) de *Streptococcus pyogenes* (véanse, por ejemplo, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLoughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); y "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012).

Debido a que las nucleasas programables por ARN (por ejemplo, Cas9) usan hibridación de ARN:ADN para dirigir sitios de escisión de ADN, estas proteínas son capaces de ser dirigidas, en principio, a cualquier secuencia especificada por el ARN guía. Se conocen en la técnica los métodos de uso de nucleasas programables por ARN, tales como Cas9, para escisión específica de sitio (por ejemplo, para modificar un genoma) (véanse, por ejemplo, Cong, L. et al. *Multiplex genome engineering using CRISPR/Cas systems*. *Science* 339, 819-823 (2013); Mali, P. et

al. RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013); Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013); Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013); Dicarlo, J.E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239 (2013)).

El término "sujeto", como se usa en el presente documento, se refiere a un organismo individual, por ejemplo, un mamífero individual. En algunas realizaciones, el sujeto es un ser humano. En algunas realizaciones, el sujeto es un mamífero no humano. En algunas realizaciones, el sujeto es un primate no humano. En algunas realizaciones, el sujeto es un roedor. En algunas realizaciones, el sujeto es una oveja, una cabra, ganado vacuno, un gato o un perro. En algunas realizaciones, el sujeto es un vertebrado, un anfibio, un reptil, un pez, un insecto, una mosca, o un nematodo. En algunas realizaciones, el sujeto es un animal de investigación. En algunas realizaciones, el sujeto está genéticamente manipulado, por ejemplo, un sujeto no humano genéticamente manipulado. El sujeto puede ser de cualquier sexo y en cualquier estadio de desarrollo.

El término "sitio diana" se refiere a una secuencia dentro de una molécula de ácido nucleico que es desaminada por una desaminasa o una proteína de fusión que comprende una desaminasa (por ejemplo, una proteína de fusión de dCas9-desaminasa proporcionada en el presente documento).

Los términos "tratamiento", "tratar" y "tratando" se refieren a una intervención clínica que tiene por finalidad aliviar, retrasar la aparición de, o inhibir el progreso de una enfermedad o trastorno, o uno o más síntomas de la misma, como se describe en el presente documento. Como se usa en el presente documento, los términos "tratamiento", "tratar" y "tratando" se refieren a una intervención clínica que tiene por finalidad aliviar, retrasar la aparición de, o inhibir el progreso de una enfermedad o trastorno, o uno o más síntomas de la misma, como se describe en el presente documento. En algunas realizaciones, el tratamiento se puede administrar después de que se hayan desarrollado uno o más síntomas y/o después de que se haya diagnosticado una enfermedad. En otras realizaciones, el tratamiento se puede administrar en ausencia de síntomas, por ejemplo, para prevenir o retrasar la aparición de un síntoma o inhibir la aparición o progresión de una enfermedad. Por ejemplo, el tratamiento se puede administrar a un individuo susceptible antes de la aparición de síntomas (por ejemplo, en vista de una historia de síntomas y/o en vista de factores de susceptibilidad genéticos u otros). El tratamiento también puede continuar después de que se hayan resuelto los síntomas, por ejemplo, para prevenir o retrasar su reaparición.

DESCRIPCIÓN DETALLADA DE CIERTAS REALIZACIONES DE LA INVENCION

Algunos aspectos de la presente divulgación proporcionan proteínas de fusión que comprenden un dominio de Cas9 que se une a un ARN guía (también denominado ARNg o ARNgu), que, a su vez, se une a la secuencia de ácidos nucleicos diana por hibridación de cadenas; y un dominio de edición de ADN, por ejemplo, un dominio de desaminasa que puede desaminar una nucleobase, tal como, por ejemplo, citidina. La desaminación de una nucleobase por una desaminasa puede conducir a una mutación puntual en el resto respectivo, que se denomina en el presente documento edición de ácidos nucleicos. Así, las proteínas de fusión que comprenden una variante o dominio de Cas9 y un dominio de edición de ADN se pueden usar para la edición dirigida de secuencias de ácidos nucleicos. Dichas proteínas de fusión son útiles para la edición dirigida de ADN *in vitro*, por ejemplo, para la generación de células o animales mutantes; para la introducción de mutaciones dirigidas, por ejemplo, para la corrección de defectos genéticos en células *ex vivo*, por ejemplo, en células obtenidas de un sujeto que son posteriormente reintroducidas en el mismo sujeto u otro sujeto; y para la introducción de mutaciones dirigidas, por ejemplo, la corrección de defectos genéticos o la introducción de mutaciones desactivantes en genes asociados a enfermedad en un sujeto. Normalmente, el dominio de Cas9 de las proteínas de fusión descritas en el presente documento no tiene ninguna actividad de nucleasa, pero en su lugar es un fragmento de Cas9 o una proteína o dominio dCas9. También se proporcionan métodos para el uso de proteínas de fusión Cas9 como se describen en el presente documento.

Se proporcionan en el presente documento dominios de Cas9 de nucleasa inactiva a modo de ejemplo no limitantes. Un dominio de Cas9 de nucleasa inactiva adecuado a modo de ejemplo es el mutante del dominio de Cas9 D10A/H840A:

MDKKYSIGLAIGTNSVGVAVITDEYKVPSSKFKVGLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
 KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDDKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDFKLF IQLVQTYNQLFEEENPINASGVDAKAIL SARLSKSRLENL
 IAQLPGEKKNGLFGNLIALSIGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKF
 IKPILKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVDK GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
 KVYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
 GKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANLAGSPA I KKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
 QELDINRLSDYDVAIVPQSF LKDDSIDNKVLTRSDKNRGS DNVPSSEVVKKMKNYWRQLLNAKLI TQRKFDNL
 TKAERGGSELKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKV
 REINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKD
 WDPKKYGGFDSPTVAYSVLVAKVEKGSKLLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISE
 FSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLEFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
 SITGLYETRIDLSQLGGD (SEQ ID NO: 37; véase, por ejemplo, *Qi et al.*, Repurposing CRISP as
 an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013;
 152(5):1173-83).

Los dominios de Cas9 de nucleasa inactiva adecuados adicionales serán evidentes para los expertos en la técnica basándose en la presente divulgación. Dichos dominios de Cas9 de nucleasa inactiva adecuados a modo de ejemplo adicionales incluyen, pero no se limitan a, dominios de mutantes D10A, D10A/D839A/H840A y D10A/D839A/H840A/N863A (véase, por ejemplo, Prashant et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838).

Proteínas de fusión entre Cas9 y enzimas o dominios de edición de ácidos nucleicos

Algunos aspectos de la presente divulgación proporcionan proteínas de fusión que comprenden (i) una enzima o dominio Cas9 de nucleasa inactiva; y (ii) una enzima o dominio de edición de ácidos nucleicos. La enzima o dominio de edición de ácidos nucleicos de la proteína de fusión de la invención comprende o es un dominio de desaminasa. En algunas realizaciones, la desaminasa es una citidina desaminasa. En algunas realizaciones, la desaminasa es una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC). En algunas realizaciones, la desaminasa es una desaminasa de la familia APOBEC1. En algunas realizaciones, la desaminasa es una citidina desaminasa inducida por activación (AID). En algunas realizaciones, la desaminasa es una ACF1/ASE desaminasa. En algunas realizaciones, la desaminasa es una adenosina desaminasa. En algunas realizaciones, la desaminasa es una desaminasa de la familia ADAT. Algunas enzimas y dominios de edición de ácidos nucleicos, así como las proteínas de fusión de Cas9 que incluyen dichas enzimas o dominios, se describen con detalle en el presente documento. Las enzimas o dominios de edición de ácidos nucleicos adecuados adicionales serán evidentes para el experto basándose en la presente divulgación.

La presente divulgación proporciona proteínas de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos de diversas configuraciones. En algunas realizaciones, la enzima o dominio de edición de ácidos nucleicos se fusiona con el extremo N del dominio de Cas9. En algunas realizaciones, la enzima o dominio de edición de ácidos nucleicos se fusiona con el extremo C del dominio de Cas9. En algunas realizaciones, el dominio de Cas9 y la enzima o dominio de edición de ácidos nucleicos se fusionan por un conector. En algunas realizaciones, el conector comprende un motivo (GGGS)_n (SEQ ID NO: 91), (G)_n, (EAAAK)_n (SEQ ID NO: 5), (GGS)_n, SGSETPGTSESATPES (SEQ ID NO: 93) (véase, por ejemplo, Guillinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82), o un motivo (XP)_n, o una combinación de cualquiera de estos, en donde n es independientemente un número entero entre 1 y 30. En algunas realizaciones, n es independientemente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30, o, si está presente más de un conector o más de un motivo de conector, cualquier combinación de los mismos. Los motivos de conector y configuraciones de conector adecuados adicionales serán evidentes para los expertos en la técnica. En algunas realizaciones, motivos y configuraciones de conector adecuados incluyen los descritos en Chen et al., Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013; 65(10):1357-69. Las secuencias conectoras adecuadas adicionales serán evidentes para los expertos en la técnica basándose en la presente divulgación.

En algunas realizaciones, la arquitectura general de las proteínas de fusión de Cas9 a modo de ejemplo proporcionadas en el presente documento comprende la estructura:

[NH₂]-[enzima o dominio de edición de ácidos nucleicos]-[Cas9]-[COOH] o

[NH₂]-[Cas9]-[enzima o dominio de edición de ácidos nucleicos]-[COOH],

5 en donde NH₂ es el extremo N de la proteína de fusión y COOH es el extremo C de la proteína de fusión.

Pueden estar presentes características adicionales, por ejemplo, una o más secuencias conectoras entre NLS y el resto de la proteína de fusión y/o entre la enzima o dominio de edición de ácidos nucleicos y Cas9. Otras características a modo de ejemplo que pueden estar presentes son las secuencias de localización, tales como secuencias de localización nuclear, secuencias de localización citoplásmica, secuencias de exportación, tales como secuencias de exportación nuclear, u otras secuencias de localización, así como marcas de secuencia que son útiles para solubilización, purificación o detección de las proteínas de fusión. Se proporcionan en el presente documento secuencias señal de localización y marcas de secuencias de proteína adecuadas, e incluyen, pero no se limitan a, 10 marcas de proteína transportadora de biotina carboxilasa (BCCP), marcas myc, marcas de calmodulina, marcas FLAG, marcas de hemaglutinina (HA), marcas de polihistidina, también denominadas marcas de histidina o marcas His, marcas de proteína de unión a maltosa (MBP), marcas nus, marcas de glutatión-S-transferasa (GST), marcas de proteína verde fluorescente (GFP), marcas de tioredoxina, marcas S, Softags (por ejemplo, Softag 1, Softag 3), marcas strep, marcas de biotina ligasa, marcas FIAH, marcas V5 y marcas SBP. Las secuencias adecuadas adicionales serán evidentes para los expertos en la técnica.

La enzima o dominio de edición de ácidos nucleicos de la proteína de fusión de la invención es una desaminasa. Por ejemplo, en algunas realizaciones, la arquitectura general de proteínas de fusión de Cas9 a modo de ejemplo con una enzima o dominio de desaminasa comprende la estructura:

[NH₂]-[NLS]-[Cas9]-[desaminasa]-[COOH],

[NH₂]-[NLS]-[desaminasa]-[Cas9]-[COOH],

[NH₂]-[Cas9]-[desaminasa]-[COOH], o

25 [NH₂]-[desaminasa]-[Cas9]-[COOH]

en donde NLS es una señal de localización nuclear, NH₂ es el extremo N de la proteína de fusión y COOH es el extremo C de la proteína de fusión. En algunas realizaciones, un conector se inserta entre Cas9 y la desaminasa. En algunas realizaciones, NLS se localiza en el extremo C de la desaminasa y/o el dominio de Cas9. En algunas realizaciones, NLS se localiza entre la desaminasa y el dominio de Cas9. También pueden estar presentes características adicionales, tales como marcas de secuencia.

Un tipo adecuado a modo de ejemplo de enzimas y dominios de edición de ácidos nucleicos son las citosina desaminasas, por ejemplo, de la familia APOBEC. La familia del complejo de edición de ARNm de apolipoproteína B (APOBEC) de enzimas citosina desaminasa engloba once proteínas que sirven para iniciar la mutagénesis en un modo controlado y beneficioso.²⁹ Un miembro de la familia, la citidina desaminasa inducida por activación (AID), es responsable de la maduración de anticuerpos que convierten citosinas en ADNmc en uracilos en un modo dependiente de la transcripción de cadenas sesgadas.³⁰ La enzima del complejo 3 de edición de apolipoproteína B (APOBEC3) proporciona protección a células humanas contra una cierta cepa del VIH-1 por la desaminación de citosinas en ADNmc viral transcrito de forma inversa.³¹ Estas proteínas requieren todas un motivo de coordinación de Zn²⁺ (His-X-Glu-X₂₃₋₂₆-Pro-Cys-X₂₋₄-Cys) y molécula de agua unida para actividad catalítica. El resto Glu actúa activando la molécula de agua para un hidróxido de cinc para el ataque nucleófilo en la reacción de desaminación. Cada miembro de la familia desamina preferencialmente en su propio "punto caliente" particular, que varía desde WRC (W es A o T, R es A o G) para hAID, hasta TTC para hAPOBEC3F.³² Una estructura cristalina reciente del dominio catalítico de APOBEC3G (Figura 2) reveló una estructura secundaria comprendida de un núcleo de lámina β de cinco cadenas flanqueado por seis hélices α, que se cree que se conservan a través de toda la familia.³³ Se ha mostrado que los bucles de centro activo son responsables de tanto la unión de ADNmc como de la determinación de la identidad de "puntos calientes".³⁴ La expresión en exceso de estas enzimas se ha asociado a inestabilidad genómica y cáncer, resaltando así la importancia del direccionamiento específico de secuencia.³⁵

Otro tipo adecuado a modo de ejemplo de enzimas y dominios de edición de ácidos nucleicos son las adenosina desaminasas. Por ejemplo, se pueden fusionar una adenosina desaminasa de la familia ADAT con un dominio de Cas9, por ejemplo, un dominio de Cas9 de nucleasa inactiva, dando así una proteína de fusión de Cas9-ADAT.

Algunos aspectos de la presente divulgación proporcionan una serie sistemática de fusiones entre Cas9 y enzimas desaminasas, por ejemplo, enzimas citosina desaminasas tales como enzimas APOBEC, o enzimas adenosina desaminasas tales como enzimas ADAT, que se han generado para dirigir las actividades enzimáticas de estas desaminasas a un sitio específico en ADN genómico. Las ventajas de uso de Cas9 como agente de reconocimiento son dobles: (1) la especificidad por secuencia de Cas9 puede ser fácilmente alterada simplemente cambiando la

secuencia de ARN_g; y (2) Cas9 se une a su secuencia diana desnaturalizando el ADN_{bc}, dando como resultado una extensión de ADN que es monocatenaria y, por tanto, un sustrato viable para la desaminasa. Se han generado proteínas de fusión satisfactorias con dominios humanos y de ratón de desaminasas, por ejemplo, dominios de AID. También se contempla una variedad de otras proteínas de fusión entre los dominios catalíticos de AID humana y de ratón y Cas9. Se entenderá que también se pueden usar otros dominios catalíticos, o dominios catalíticos de otras desaminasas, para generar proteínas de fusión con Cas9, y que la divulgación no está limitada a este respecto.

En algunas realizaciones, se proporcionan proteínas de fusión de Cas9 y AID. En un esfuerzo por manipular las proteínas de fusión de Cas9 para aumentar las velocidades de mutación en ADN_{mc}, se unieron tanto AID de ratón como humano al gen V de fago filamentoso (una proteína de unión de ADN_{mc} no específico). Las proteínas de fusión resultantes presentaron actividades mutagénicas potenciadas en comparación con las enzimas no mutantes en un ensayo basado en células. Este trabajo demuestra que la actividad enzimática de estas proteínas se mantiene y puede ser satisfactoriamente dirigido a secuencias genéticas con proteínas de fusión.³⁶

Aunque se ha informado de varias estructuras cristalinas de Cas9 (e incluso Cas9 en complejo con su ARN_g y ADN diana) (véanse, por ejemplo, Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014; 343(6176):1247997. PMID: 24505130; y Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014; 156(5):935-49. PMID: 24529477), es desconocida la porción de ADN que es monocatenaria en el complejo Cas9-ADN (el tamaño de la burbuja de Cas9-ADN). Sin embargo, se ha mostrado en un sistema de dCas9 con un ARN_g específicamente diseñado para el complejo para interferir con transcripción que la interferencia transcripcional solo ocurre cuando el ARN_g se une a la cadena no de molde. Este resultado sugiere que ciertas porciones del ADN en el complejo de ADN-Cas9 son expuestas por el Cas9, y podrían ser posiblemente dirigidas por una desaminasa en la proteína de fusión (véase Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152(5):1173-83. PMID: 23452860). Soportando además está noción, experimentos de reconocimiento con exonucleasa III y nucleasa P1 (que solo actúa en ADN_{mc} como sustrato) han revelado que al menos 26 bases en la cadena de no molde son susceptibles a digestión por estas enzimas (véase Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014; 343(6176):1247997. PMID: 24505130). También se ha informado que en ciertos casos, Cas9 induce mutaciones de sustitución de bases individuales en esta extensión susceptible de ADN a frecuencias de hasta 15 % (véase Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*. 2014; 32(6):569-76. PMID: 24770325). Aunque el mecanismo de introducción de estas mutaciones es desconocido, en todos los casos, la base que se muta es una citosina, que podría indicar posiblemente la participación de una enzima citosina desaminasa. Tomados conjuntamente, estos datos están claramente de acuerdo con una porción del ADN diana que es monocatenario y susceptible a otras enzimas. Se ha mostrado en un sistema de dCas9 con un ARN_g específicamente diseñado para el complejo para interferir con la transcripción que la interferencia transcripcional solo ocurre cuando el ARN_g se une a la cadena no de molde. Este resultado sugiere que ciertas porciones del ADN en el complejo de ADN-Cas9 son expuestas por el Cas9, y podrían ser posiblemente dirigidas por AID en la proteína de fusión.¹⁶ Por consiguiente, tanto las fusiones del extremo N como del extremo C de Cas9 con un dominio de desaminasa son útiles según aspectos de la presente divulgación.

En algunas realizaciones, el dominio de desaminasa y el dominio de Cas9 se fusionan entre sí por un conector. Se pueden emplear diversas longitudes y flexibilidades de conector entre el dominio de desaminasa (por ejemplo, AID) y el dominio de Cas9 (por ejemplo, que varía desde conectores muy flexibles de la forma (GGGGG)_n (SEQ ID NO: 91), (GGG)_n y (G)_n hasta conectores más rígidos de la forma (EAAAK)_n (SEQ ID NO: 5), SGSETPGTSESATPES (SEQ ID NO: 93) (véase, por ejemplo, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol*. 2014; 32(6): 577-82) y (XP)_n³⁷ para lograr la longitud óptima para la actividad de desaminasa para la aplicación específica.

Se proporcionan a continuación algunas enzimas y dominios de edición de ácidos nucleicos adecuados a modo de ejemplo, por ejemplo, desaminasas y dominios de desaminasa, que se pueden fusionar a dominios de Cas9 según aspectos de la presente divulgación. Se entenderá que, en algunas realizaciones, se puede usar el dominio activo de la secuencia respectiva, por ejemplo, el dominio sin una señal de localización (señal de localización nuclear, sin señal de exportación nuclear, señal de localización citoplásmica).

AID humana:

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKC
 HVELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
 LYFCEDRKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
 SVRLSRQLRRILLPLYEVDDLRLDAFRTLGL (SEQ ID NO: 6)

(subrayado: señal de localización nuclear; subrayado doble: señal de exportación nuclear)

AID de ratón:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTAR
LYFCEDRKAPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHEN
SVRLTRQLRRILLPLYEVDDL RDAFRMLGE (SEQ ID NO: 7)

(subrayado: señal de localización nuclear; subrayado doble: señal de exportación nuclear)

5 AID de perro:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR
LYFCEDRKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHEN
SVRLSRQLRRILLPLYEVDDL RDAFRTLGL (SEQ ID NO: 8)

(subrayado: señal de localización nuclear; subrayado doble: señal de exportación nuclear)

AID bovino:

MDSLLKKQRFQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTAR
LYFCDKERKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHE
NSVRLSRQLRRILLPLYEVDDL RDAFRTLGL (SEQ ID NO: 9)

10

(subrayado: señal de localización nuclear; subrayado doble: señal de exportación nuclear)

APOBEC-3 de ratón:

MGPFCLGCSHRKCYSPIRNLISQETFKHFKNLGYAKGRKDTFLCYEVTRKDCDSPV
SLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFL*
ATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDN
GGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVSSSSSTLSNICLTKGLPETRFCVEG
RRMDPLSEEEFYQFYQNRVKHLCYHRMKPYLCYQLEQFNGQAPLKGCLLSEK GK
QHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHW
KRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRL
RRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 10)

(cursiva: dominio de edición de ácidos nucleicos)

15

APOBEC-3 de rata:

MGPFC^LGC^SHRKCYSP^IRNLISQ^ETFKF^HFKNLRYAID^RKDTFLCYEV^TRKDC^DSPVS
 LHHGV^FKNKDNI^HAEIC^FLYWF^HDKVL^KVLS^PREE^FKITWYMS^WSP^CFECAEQVLR^FLA
 THHNLS^LDIFSS^RLYNIRD^PENQQNL^CRLVQ^EGAQVAAMD^LYEFK^KCW^KKFVD^NGG
 RRF^RPW^KKLLTN^FRYQ^DSKLQ^EILR^PCYIP^VPSS^SSTLS^NIC^LT^KGL^PETR^FCV^ERRR
 VHLL^SEEEF^YSQ^FYNQR^VKHLC^YYHG^VKPY^LCYQ^LE^QFNGQ^APL^KG^CLL^SE^KG^KQH
 AEIL^FLD^KIR^SMEL^SQ^VIIT^CYLT^WSP^CPN^CAW^QLA^AFK^RDR^PDL^LHI^YTS^RLY^FHW^KR
 PFQ^KGL^CSL^WQ^SGIL^VDM^LPQ^FTD^CWT^NFV^NPK^RPF^WPW^KGLE^IIS^RRT^QRR^LHR
 IKES^WGL^QDL^VNDF^GNL^QLG^PPMS (SEQ ID NO: 11)

(cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3G de macaco rhesus:

MVEPMD^PRT^FVSN^FNN^RPIL^SGL^NT^VW^LCCE^VKTK^DPS^GP^PLD^AKIF^QG^KV^YSK^AK^Y
HP^EMR^FLR^WFK^WR^QL^HHD^QE^YK^VT^WY^VSW^SP^CTR^CANS^VAT^FLA^KDP^KV^TL^TIF^VA
 RLY^YFW^KPD^YQ^QAL^RIL^CQ^KR^GGP^HAT^MK^IM^NY^NE^FQ^DC^WN^KF^VD^GR^GK^PFK^PR^N
 5 NLP^KHY^TLL^QAT^LGEL^LR^HLM^DPG^TFT^SN^FNN^KP^WV^SG^QHET^YLC^YK^VER^LH^NDT
 WV^PLN^QHR^GFL^RN^QAP^NIHG^FPK^RHA^ELC^FLD^LIP^FW^KLD^GQ^QY^RVT^CFT^SW^SP^CF^S
 CA^QE^MA^KF^IS^NNE^HV^SLC^IFA^ARI^YDD^QGR^YQ^EGL^RAL^HRD^GAK^IAM^MN^YSE^FE^YC
 W^DT^FV^DR^QGR^PF^QP^WD^GL^DE^HS^QAL^SG^RL^RAI (SEQ ID NO: 12)

(cursiva: dominio de edición de ácidos nucleicos; subrayado: señal de localización citoplásmica)

APOBEC-3G de chimpancé:

MK^PHR^NP^VER^MY^QDT^FSD^NF^YN^RPIL^SHR^NT^VW^LCY^EV^KTK^GPS^RP^PLD^AKIF^RG^Q
V^YSK^LK^YHP^EMR^FFF^HW^FSK^WR^KL^HRD^QE^YE^VT^WY^ISW^SP^CTK^CTR^DV^AT^FLA^EDP^KV
 TLT^IF^VAR^LY^YF^WDP^YQ^EAL^RSL^CQ^KRD^GP^RAT^MK^IM^NY^DE^FQ^HC^WS^KF^VY^SQ^RE
 LF^EP^WNN^LP^KY^YILL^HIM^LGE^ILR^HSM^DP^PT^FT^SN^FNN^LW^VR^GR^HE^TY^LC^YE^VER^L
 H^NDT^WV^{LL}N^QRR^GFL^CN^QAP^HK^HG^FLE^GR^HAE^LC^FLD^VIP^FW^KLD^LH^QD^YR^VT^CFT^S
 W^SP^CF^SCA^QE^MA^KF^IS^NN^KH^VSL^CI^FA^ARI^YDD^QGR^CQ^EGL^RT^LAK^AG^AK^IS^IM^TY^SE
 FK^HC^WD^TF^VD^HQ^GCP^FQ^PW^DGLE^EHS^QAL^SG^RL^RAIL^QN^QGN (SEQ ID NO: 13)

10 (cursiva: dominio de edición de ácidos nucleicos; subrayado: señal de localización citoplásmica)

APOBEC-3G de mono verde:

MN^PQ^IR^NM^VE^QM^EP^DIF^VY^FNN^RPIL^SGR^NT^VW^LCY^EV^KTK^DPS^GP^PLD^ANIF^QG^K
LY^EAK^DHP^EM^KFL^HW^FR^KWR^QL^HRD^QE^YE^VT^WY^VSW^SP^CTR^CANS^VAT^FLA^EDP^KV
 TLT^IF^VAR^LY^YF^WK^PDY^QQ^QAL^RIL^CQ^ER^GGP^HAT^MK^IM^NY^NE^FQ^HC^WN^EF^VD^GQ^G
 K^PFK^PR^KN^LP^KHY^TLL^HAT^LGEL^LR^HV^MD^PGT^FT^SN^FNN^KP^WV^SG^QR^ET^YLC^YK^VE
 R^SH^NDT^WV^{LL}N^QHR^GFL^RN^QAP^DR^HG^FPK^RHA^ELC^FLD^LIP^FW^KLD^DQ^QY^RVT^CFT^S
 S^WP^CF^SCA^QK^MA^KF^IS^NN^KH^VSL^CI^FA^ARI^YDD^QGR^CQ^EGL^RT^LHR^DG^AK^IA^VM^NY
 SE^FE^YC^WD^TF^VD^RQ^{GR}P^FQ^PW^DGLE^HS^QAL^SG^RL^RAI (SEQ ID NO: 14)

(cursiva: dominio de edición de ácidos nucleicos; subrayado: señal de localización citoplásmica)

APOBEC-3G humana:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPLDAKIFRGQ
VYSELKYHPEMRFFHWF~~SKWRKLHRDQEYEV~~TWYISWSPCTKCTRDMATFLAEDPKV
TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE
LFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNNEPWVRGRHETYLCYEVERM
HN~~DTWVLLNQR~~RGF~~LCNQAPHKHGFLEGR~~HAELCFLDVIPFWKLDLDQDYRVTCFTS
WSPCFSQAQEMAKFISK~~NKHVSLCIFTARIYDDQGR~~CQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQCPFPQWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 15)

(cursiva: dominio de edición de ácidos nucleicos; subrayado: señal de localización citoplásmica)

APOBEC-3F humana:

5 MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPLDAKIFRGQ
VYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAHPNVTL
TISAARLYYYWERDYRRALCRLSQA~~GARVKIMDDEEFAYC~~WENFVYSEGQPFMPW
YKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVK
HHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECA
GEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCW
ENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 16)

(cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3B humana:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPN
VTLTISAARLYYYWERDYRRALCRLSQA~~GARVTIMDYEEFAYC~~WENFVYNEGQQF
MPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLD
NGTWVLMQHMGLFCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWS
PCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTY
DEFEYCWDTFVYRQGPCFPQWDGLEEHSQALS~~SGRLRAILQNQGN~~ (SEQ ID NO: 17)

10 (cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3C humana:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVF
RNQVDSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSN
VNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKP
WKGLKTNFRLKRRRLRESLQ (SEQ ID NO: 18)

(cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3A humana:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLH
NQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFWSWGCAGEVRAFLQ
ENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGC
PFQPWDGLDEHSQALSGRRLRAILQNQGN (SEQ ID NO: 19)

(cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3H humana:

MALLTAETFRLLQFNKRRLRRPYYPKALLCYQLTPQNGSTPTRGYFENKKKCHAEI
CFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWC
KPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNRA
IKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 20)

5

(cursiva: dominio de edición de ácidos nucleicos)

APOBEC-3D humana:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFWVSWNPCLPCVV
KVTKFLAEHPNVTLTISAARLYYYRDRDWRVLLRLHKAGARVKIMDYEDFAYCW
ENFVCNEGQPFMPWYKFDNDYASLHRTLKEILRNPMEAMYPHIFYPHFKNLLKACG
RNESWLCFTMEVTKHHSVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNY
EVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSSLSQEGAS
VKIMGYKDFVSCWKNFVYSDDPEFKPWKGLQTNFRLKRRRLREILQ (SEQ ID NO:
21)

(cursiva: dominio de edición de ácidos nucleicos)

10 APOBEC-1 humana:

MTSEKGPSTGDPTLRRRIEPWFDVVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKN
TTNHVEVNFIIKFTSERDFHPSMSCSITWFLSWSPCWECQAIREFLSRHPGVTLVIYV
ARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQY
PPLWMMLYALELHCILSLPPCLKISRWRQNHLTFFRLHLQNCHYQTIPPHILLATGLI
HPSVAWR (SEQ ID NO: 22)

APOBEC-1 de ratón:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQN
TSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYA
RLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHL
WVKLYVLELYCILGLPPCLKILRRKQPQLTFFTTITLQTCHYQRIPPHLLWATGLK
(SEQ ID NO: 23)

APOBEC-1 de rata:

MSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNEAHWPRYPHLW
VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ
ID NO: 24)

ADAT-2 humana:

MEAKAAPKPAASGACSVSAEETEKWMEEAMHMAKEALENTEVPVGCLMVYNNEV
VGKGRNEVNQTKNATRHAEMVAIDQVLDWCRQSGKSPSEVFEHTVLYVTVEPCIM
CAAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGRPFQCIPGYRAEEAVEM
LKTFYKQENPNAPKSKVRKKECQKS (SEQ ID NO: 25)

5 ADAT-2 de ratón:

MEEKVESTTTTDPGPCVVSQETEKWMEEAMRMAKEALENIEVPVGCLMVYNNEVV
GKGRNEVNQTKNATRHAEMVAIDQVLDWCHQHGGQSPSTVFEHTVLYVTVEPCIMC
AAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGRPFQCIPGYRAEEAVELL
KTFYKQENPNAPKSKVRKKDCQKS (SEQ ID NO: 26)

ADAT-1 de ratón:

MWTADEIAQLCYAHYNVRLPKQKGPEPNREWTLAAVVKIQASANQACDIPEKEVQ
VTKEVVSMGTGTCIGQSKMRESGDILNDSHAEIARRSFQRYLLHQLHLAAVLKEDSIFV
PGTQRLWRLRPDLFSVFFSSHTPCGDASHIPMLEFEEQPCCPVIRSWANNSPVQETENLE
DSKDKRNCEDPASPVAKKMRLGTPARSLSNCAHHGTQESGPVKPDVSSDLTKKEEPDAA
NGIASGSFRVVDVYRTGAKCVPGETGDLREPGAAYHQVGLLRVVKPGRGDRTCSMSCSDK
MARWNVLGCQGALLMHFLEKPIYLSAVVIGKCPYSQEAMRRALTGRCEETLVLPGRFGVQ
ELEIQQSGLLFEQSRCAVHRKRGDSPGRLVPCGAAISWSAVPQQPLDVTANGFPQGTTKK
EIGSPRARSRISKVELFRSFQKLLSSIADDEQPDSSIRVTKKLDTYQEYKDAASAYQEAWGAL
RRIQPFASWIRNPPDYHQFK (SEQ ID NO: 27)

(cursiva: dominio de edición de ácidos nucleicos)

10 ADAT-1 humana:

MWTADEIAQLCYEHYGIPLPKKKGPEPNHEWTLAAVVKIQSPADKACDTPDKPVQ
VTKEVVSMGTGTCIGQSKMRKNGDILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIF
VPGTQKGVWKLRRDLIFVFFSSHTPCGDASHIPMLEFEDQPCCPVFRNWAHNSSEASSNL
EAPGNERKCEDPDSVTKMRLEPGTAAREVTNGAAHHQSFQKQSGPISPGIHSCLTV
EGLATVTRIPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFHQVGLLRVVKPGRGDRTRSM
SCSDKMARWNVLGCQGALLMHLEEPIYLSAVVIGKCPYSQEAMQALIGRCQNVSALPK
GFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQ
GTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDKWP HSLRVQKLDTYQEYKEAASSYQEA
WSTLRKQVFGSWIRNPPDYHQFK (SEQ ID NO: 28)

(cursiva: dominio de edición de ácidos nucleicos)

En algunas realizaciones, las proteínas de fusión como se proporcionan en el presente documento comprenden el aminoácido de longitud completa de una enzima de edición de ácidos nucleicos, por ejemplo, una de las secuencias proporcionadas anteriormente. En otras realizaciones, sin embargo, las proteínas de fusión como se proporcionan en el presente documento no comprenden una secuencia de longitud completa de una enzima de edición de ácidos nucleicos, sino solo un fragmento de la misma. Por ejemplo, en algunas realizaciones, una proteína de fusión proporcionada en el presente documento comprende un dominio de Cas9 y un fragmento de una enzima de edición de ácidos nucleicos, por ejemplo, en donde el fragmento comprende un dominio de edición de ácidos nucleicos. Las secuencias de aminoácidos a modo de ejemplo de dominios de edición de ácidos nucleicos se muestran en las secuencias anteriores como letras en cursiva, y secuencias adecuadas adicionales de dichos dominios serán evidentes para los expertos en la técnica.

Las secuencias de enzimas de edición de ácidos nucleicos adecuadas adicionales, por ejemplo, enzima desaminasa y secuencias de dominio, que se pueden usar según aspectos de la presente invención, por ejemplo, que se pueden fusionar con un dominio de Cas9 de nucleasa inactiva, serán evidentes para los expertos en la técnica basándose en la presente divulgación. En algunas realizaciones, dichas secuencias de enzima adicionales incluyen enzima desaminasa o secuencias de dominio de desaminasa que son al menos 70 %, al menos 75 %, al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, o al menos 99 % similares a las secuencias proporcionadas en el presente documento. Dominios, variantes y secuencias de Cas9 adecuados adicionales también serán evidentes para los expertos en la técnica. Los ejemplos de dichos dominios de Cas9 adecuados adicionales incluyen, pero no se limitan a, dominios de mutantes D10A, D10A/D839A/H840A y D10A/D839A/H840A/N863A (véase, por ejemplo, Prashant et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838).

Las estrategias adecuadas adicionales para generar proteínas de fusión que comprenden un dominio de Cas9 y un dominio de desaminasa serán evidentes para los expertos en la técnica basándose en la presente divulgación en combinación con el conocimiento general en la técnica. Las estrategias adecuadas para generar proteínas de fusión según aspectos de la presente divulgación usando conectores o sin el uso de conectores también serán evidentes para los expertos en la técnica en vista de la presente divulgación y el conocimiento en la técnica. Por ejemplo, Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2):442-51, mostraron que las fusiones de extremo C de Cas9 con VP64 usando 2 NLS como conector (SPKKKRVKVEAS, SEQ ID NO: 29) se pueden emplear para activación transcripcional. Mali et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013; 31(9):833-8, informaron que las fusiones de extremo C con VP64 sin conector se pueden emplear para activación transcripcional. Y Maeder et al., CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013; 10: 977-979, informaron que las fusiones de extremo C con VP64 usando un conector Gly₄Ser (SEQ ID NO: 91) se pueden usar como activadores transcripcionales. Recientemente, se han generado satisfactoriamente fusiones de nucleasa dCas9-FokI y presentan especificidad enzimática mejorada en comparación con la enzima Cas9 parental (en Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol*. 2014; 32(6): 577-82, y en Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*. 2014; 32(6):569-76. PMID: 24770325 Se usó un conector SGSETPGTSESATPES (SEQ ID NO: 93) o GGGGS (SEQ ID NO: 91) en proteínas de fusión de FokI-dCas9, respectivamente).

Uso de proteínas de fusión de edición de ADN de Cas9 para corregir mutaciones asociadas a enfermedad

Algunas realizaciones proporcionan métodos para usar las proteínas de fusión de edición de ADN de Cas9 proporcionadas en el presente documento. En algunas realizaciones, la proteína de fusión se usa para introducir una mutación puntual en un ácido nucleico por desaminación de una nucleobase diana, por ejemplo, un resto C. En algunas realizaciones, la desaminación de la nucleobase diana da como resultado la corrección de un defecto genético, por ejemplo, en la corrección de una mutación puntual que conduce a una pérdida de función en un producto génico. En algunas realizaciones, el defecto genético está asociado con una enfermedad o trastorno, por ejemplo, un trastorno de almacenamiento lisosómico o una enfermedad metabólica, tal como, por ejemplo, diabetes de tipo I. En algunas realizaciones, los métodos proporcionados en el presente documento se usan para introducir una mutación puntual desactivante en un gen o alelo que codifica un producto génico que está asociado con una enfermedad o trastorno. Por ejemplo, en algunas realizaciones, se proporcionan métodos en el presente documento que emplean una proteína de fusión de edición de ADN de Cas9 para introducir una mutación puntual desactivante en un oncogén (por ejemplo, en el tratamiento de una enfermedad proliferativa). Una mutación desactivante puede generar, en algunas realizaciones, un codón de terminación prematuro en una secuencia codificante, que da como resultado la expresión de un producto génico truncado, por ejemplo, una proteína truncada que carece de la función de la proteína de longitud completa.

En algunas realizaciones, el fin de los métodos proporcionados en el presente documento es restaurar la función de un gen disfuncional mediante edición del genoma. Las proteínas de fusión de Cas9-desaminasa proporcionadas en el presente documento pueden ser validadas para agentes terapéuticos humanos basados en edición génica *in vitro*,

por ejemplo, corrigiendo una mutación asociada a enfermedad en cultivo celular humano. Se entenderá por el experto que las proteínas de fusión proporcionadas en el presente documento, por ejemplo, las proteínas de fusión que comprenden un dominio de Cas9 y un dominio de desaminasa de ácidos nucleicos se pueden usar para corregir cualquier mutación puntual individual T -> C o A -> G. En el primer caso, la desaminación de C mutante de nuevo a U corrige la mutación, y en el último caso, la desaminación de C que está apareada con base con G mutante, seguido por una ronda de replicación, corrige la mutación.

Una mutación relevante de enfermedad a modo de ejemplo que se puede corregir por las proteínas de fusión proporcionadas *in vitro* o *in vivo* es el polimorfismo H1047R (A3140G) en la proteína PI3KCA. La fosfoinositida-3-cinasa, proteína de subunidad alfa catalítica (PI3KCA), actúa fosforilando el grupo 3-OH del anillo inositol de fosfatidilinositol. Se ha encontrado que el gen PI3KCA se muta en muchos carcinomas diferentes, y así se considera que es un potente oncogén.⁵⁰ En realidad, la mutación A3140G está presente en varias líneas NCI-60 de células cancerosas, tales como, por ejemplo, las líneas celulares HCT116, SKOV3 y T47D, que están fácilmente disponibles de la Colección Americana de Cultivos Tipo (ATCC).⁵¹

En algunas realizaciones, una célula que lleva una mutación a corregir, por ejemplo, una célula que lleva una mutación puntual, por ejemplo, una mutación puntual A3140G en el exón 20 del gen PI3KCA, dando como resultado una sustitución H1047R en la proteína PI3KCA, se pone en contacto con una construcción de expresión que codifica una proteína de fusión de Cas9-desaminasa y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación respectivo en el gen PI3KCA codificante. Se pueden realizar experimentos de control donde el ARN_g se diseña para dirigir las enzimas de fusión a restos distintos de C que están dentro del gen PI3KCA. Se puede extraer ADN genómico de las células tratadas, y amplificar y secuenciar la secuencia relevante de la PCR de genes PI3KCA para evaluar las actividades de las proteínas de fusión en cultivo de células humanas.

Se entenderá que el ejemplo de corregir mutaciones puntuales en PI3KCA se proporciona para fines de ilustración y no pretende limitar la presente divulgación. El experto entenderá que las proteínas de fusión de edición de ADN presentemente desveladas se pueden usar para corregir otras mutaciones puntuales y mutaciones asociadas a otros cánceres y con enfermedades distintas de cáncer que incluyen otras enfermedades proliferativas.

La satisfactoria corrección de mutaciones puntuales en genes y alelos asociados a enfermedad facilita nuevas estrategias de corrección génica con aplicaciones en agentes terapéuticos e investigación básica. Los sistemas de modificación de bases individuales específicos de sitio como las fusiones desveladas de Cas9 y enzimas o dominios desaminasa también tienen aplicaciones en terapia génica "inversa", donde ciertas funciones génicas son deliberadamente suprimidas o abolidas. En estos casos, se pueden usar restos Trp (TGG), Gln (CAA y CAG), o Arg (CGA) que mutan de manera específica de sitio a codones de terminación prematuros (TAA, TAG, TGA) para abolir la función de proteínas *in vitro*, *ex vivo*, o *in vivo*.

Las proteínas de fusión de la presente invención se pueden usar en métodos para el tratamiento de un sujeto diagnosticado con una enfermedad asociada a o provocada por una mutación puntual que se puede corregir por una proteína de fusión de edición de ADN de Cas9 proporcionada en el presente documento. Por ejemplo, en algunas realizaciones, el método comprende administrar a un sujeto que tiene dicha enfermedad, por ejemplo, un cáncer asociado a una mutación puntual de PI3KCA como se ha descrito anteriormente, una cantidad eficaz de una proteína de fusión de Cas9-desaminasa que corrige la mutación puntual o introduce una mutación desactivante en el gen asociado a enfermedad. En algunas realizaciones, la enfermedad es una enfermedad proliferativa. En algunas realizaciones, la enfermedad es una enfermedad genética. En algunas realizaciones, la enfermedad es una enfermedad neoplásica. En algunas realizaciones, la enfermedad es una enfermedad metabólica. En algunas realizaciones, la enfermedad es una enfermedad de almacenamiento lisosómico. Otras enfermedades que se pueden tratar corrigiendo una mutación puntual o introduciendo una mutación desactivante en un gen asociado a enfermedad serán conocidas por los expertos en la técnica, y la divulgación no está limitada a este respecto.

Las proteínas de fusión de la presente invención se pueden usar en métodos para el tratamiento de enfermedades o trastornos adicionales, por ejemplo, enfermedades o trastornos que están asociados o provocados por una mutación puntual que se puede corregir por edición de genes mediados por desaminasa. Algunas de tales enfermedades se describen en el presente documento, y enfermedades adecuadas adicionales que se pueden tratar con las estrategias y proteínas de fusión proporcionadas en el presente documento serán evidentes para los expertos en la técnica basándose en la presente divulgación. Las enfermedades y trastornos adecuados a modo de ejemplo se enumeran a continuación. Se entenderá que la numeración de las posiciones o restos específicos en las secuencias respectivas depende de la proteína particular y el esquema de numeración usado. La numeración podría ser diferente, por ejemplo, en precursores de una proteína madura y la propia proteína madura, y diferencias en secuencias de especie a especie pueden afectar la numeración. Un experto en la técnica será capaz de identificar el resto respectivo en cualquier proteína homóloga y en el ácido nucleico codificante respectivo por métodos bien conocidos en la técnica, por ejemplo, por alineamiento de secuencias y determinación de restos homólogos. Las enfermedades y trastornos adecuados a modo de ejemplo incluyen, sin limitación, fibrosis quística (véanse, por ejemplo, Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell*. 2013; 13: 653-658; y Wu et al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell*. 2013; 13: 659-662, ninguno de los cuales usa una proteína de fusión de desaminasa para corregir el defecto genético); fenilcetonuria - por ejemplo, mutación de fenilalanina a serina en la posición 835

(ratón) o 240 (humano) o un resto homólogo en el gen fenilalanina hidroxilasa (mutación T>C) - véase, por ejemplo, McDonald et al., *Genomics*. 1997; 39:402-405; síndrome de Bernard-Soulier (BSS) - por ejemplo, mutación de fenilalanina a serina en la posición 55 o un resto homólogo, o cisteína a arginina en el resto 24 o un resto homólogo en la glucoproteína IX de la membrana plaquetaria (mutación T>C) - véase, por ejemplo, Noris et al., *British Journal of Haematology*. 1997; 97: 312-320, y Ali et al., *Hematol*. 2014; 93: 381-384; hiperqueratosis epidermolítica (EHK) - por ejemplo, mutación de leucina a prolina en la posición 160 o 161 (si se cuenta la metionina iniciadora) o un resto homólogo en queratina 1 (mutación T>C) - véase, por ejemplo, Chipev et al., *Cell*. 1992; 70: 821-828, véase también el número de acceso P04264 en la base de datos UNIPROT en www.uniprot.org; enfermedad pulmonar obstructiva crónica (EPOC) - por ejemplo, mutación de leucina a prolina en la posición 54 o 55 (si se cuenta la metionina iniciadora) o un resto homólogo en la forma procesada de α_1 -antitripsina o resto 78 en la forma no procesada o un resto homólogo (mutación T>C) - véase, por ejemplo, Poller et al., *Genomics*. 1993; 17: 740-743, véase también el número de acceso P01011 en la base de datos UNIPROT; enfermedad de Charcot-Marie-Toot tipo 4J - por ejemplo, mutación de isoleucina a treonina en la posición 41 o un resto homólogo en FIG4 (mutación T>C) - véase, por ejemplo, Lenk et al., *PLoS Genetics*. 2011; 7: e1002104; neuroblastoma (NB) - por ejemplo, mutación de leucina a prolina en la posición 197 o un resto homólogo en caspasa-9 (mutación T>C) - véase, por ejemplo, Kundu et al., *3 Biotech*. 2013, 3:225-234; enfermedad de von Willebrand (vWD) - por ejemplo, mutación de cisteína a arginina en la posición 509 o un resto homólogo en la forma procesada de factor de von Willebrand, o en la posición 1272 o un resto homólogo en la forma no procesada de factor de von Willebrand (mutación T>C) - véase, por ejemplo, Lavergne et al., *Br. J. Haematol*. 1992, véase también el número de acceso P04275 en la base de datos UNIPROT; 82: 66-72; miotonía congénita - por ejemplo, mutación de cisteína a arginina en la posición 277 o un resto homólogo en el gel del canal de cloruro muscular CLCN1 (mutación T>C) - véase, por ejemplo, Weinberger et al., *The J. of Physiology*. 2012; 590: 3449-3464; amiloidosis renal hereditaria - por ejemplo, mutación de codón de terminación a arginina en la posición 78 o un resto homólogo en la forma procesada de apolipoproteína All o en la posición 101 o un resto homólogo en la forma no procesada (mutación T>C) - véase, por ejemplo, Yazaki et al., *Kidney Int*. 2003; 64: 11-16; cardiomiopatía dilatada (DCM) - por ejemplo, mutación de triptófano a arginina en la posición 148 o un resto homólogo en el gen FOXD4 (mutación T>C), véase, por ejemplo, Minoretti et al., *Int. J. of Mol. Med*. 2007; 19: 369-372; linfedema hereditario - por ejemplo, mutación de histidina a arginina en la posición 1035 o un resto homólogo en tirosina cinasa de VEGFR3 (mutación A>G), véase, por ejemplo, Irrthum et al., *Am. J. Hum. Genet*. 2000; 67: 295-301; enfermedad de Alzheimer familiar - por ejemplo, mutación de isoleucina a valina en la posición 143 o un resto homólogo en presenilina 1 (mutación A>G), véase, por ejemplo, Gallo et al., *J. Alzheimer's disease*. 2011; 25: 425-431; enfermedad priónica - por ejemplo, mutación de metionina a valina en la posición 129 o un resto homólogo en la proteína priónica (mutación A>G) - véase, por ejemplo, Lewis et al., *J. of General Virology*. 2006; 87: 2443-2449; síndrome crónico, infantil, neurológico, cutáneo, articular (CINCA) - por ejemplo, mutación de tirosina a cisteína en la posición 570 o un resto homólogo en criopirina (mutación A>G) - véase, por ejemplo, Fujisawa et al. *Blood*. 2007; 109: 2903-2911; y miopatía relacionada con desmina (DRM) - por ejemplo, mutación de arginina a glicina en la posición 120 o un resto homólogo en α B cristalina (mutación A>G) - véase, por ejemplo, Kumar et al., *J. Biol. Chem*. 1999; 274: 24137-24141.

Será evidente para los expertos en la técnica que para dirigir una proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos como se desvela en el presente documento a un sitio diana, por ejemplo, un sitio que comprende una mutación puntual a editar, normalmente es necesario co-expresar la proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos junto con un ARN guía, por ejemplo, un ARN_g. Como se explica con más detalle en cualquier parte en el presente documento, un ARN guía normalmente comprende una región estructural de ARN_{cr} que permite la unión de Cas9, y una secuencia guía, que confiere especificidad de secuencia a la proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos. En algunas realizaciones, el ARN guía comprende una estructura 5'-[secuencia guía]-guuuuagagcuagaaauagcaaguuaaaauaaagguaguccguuaaucaacuugaaaaagggcaccgagucgggcuuuuu-3' (SEQ ID NO: 38), en donde la secuencia guía comprende una secuencia que es complementaria a la secuencia diana. La secuencia guía normalmente tiene 20 nucleótidos de longitud. Las secuencias de ARNs guía adecuados para dirigir las proteínas de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos a sitios diana genómicos específicos serán evidentes para los expertos en la técnica basándose en la presente divulgación. Dichas secuencias de ARN guía adecuadas normalmente comprenden secuencias guía que son complementarias a una secuencia nucleica dentro de 50 nucleótidos en la dirección 5' o en la dirección 3' del nucleótido diana a editar. Algunas secuencias guía de ARN adecuadas a modo de ejemplo para dirigir proteínas de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos a secuencias diana específicas se proporcionan a continuación.

Polimorfismo H1047R (A3140G) en la subunidad alfa catalítica de fosfoinositida-3-cinasa (PI3KCA o PIK3CA) (están subrayados la posición del nucleótido mutado y el codón respectivo):

gatgacattgcatacatttcgaaagaccctagccttagataaaaactgagcaagaggctttg
 D D I A Y I R K T L A L D K T E Q E A L
 gagtatttcatgaaacaaatgaatgatgcacgtcatggtggctggacaacaaaaatggat
 E Y F M K Q M N D A R H G G W T T K M D
 tggatcttccacacaattaacagcatgcattgaaactgaaagataaactgagaaaaatgaaa
 W I F H T I K Q H A L N - K I T E K M K

(Secuencia de nucleótidos - SEQ ID NO: 39; secuencia de proteínas - SEQ ID NO: 40).

5 Las secuencias guía adecuadas a modo de ejemplo para dirigir una proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos al resto A3140G mutante incluyen, sin limitación: 5'-aucggauctauuuugacuc-3' (SEQ ID NO: 41); 5'-ucggaucuaauuuugacucg-3' (SEQ ID NO: 42); 5'-cuuagauaaaacugagcaag-3' (SEQ ID NO: 43); 5'-aucuaauuuugacucguucuc-3' (SEQ ID NO: 44); 5'-uaaacugagcaagaggcuu-3' (SEQ ID NO: 45); 5'-ugguggcuggacaacaaaaa-3' (SEQ ID NO: 46); 5'-gcuggacaacaaaauggau-3' (SEQ ID NO: 47); 5'-guguuaauuuugucguacgua-3' (SEQ ID NO: 48). Las secuencias guía adecuadas adicionales para dirigir una proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos a una secuencia de PI3KCA mutante, a cualquiera de las secuencias adicionales proporcionadas a continuación, o a secuencias mutantes adicionales asociadas a una enfermedad, serán evidentes para los expertos en la técnica basándose en la presente divulgación.

10 Fenilcetonuria - mutación de fenilalanina a serina en el resto 240 en el gen fenilalanina hidroxilasa (mutación T>C) (están subrayadas la posición del nucleótido mutado y el codón respectivo):

aatcacatttttccacttcttgaaaagtactgtggcttccatgaagataaacattccccag
 N H I F P L L E K Y C G F H E D N I P Q
 ctggaagacgtttctcaattcctgcagacttgcactggtctccgcctccgacctgtggct
 L E D V S Q F L Q T C T G S R L R P V A
 ggctgctttcctctcgggatttcttgggtggcctggccttccgagcttccactgcaca
 G L L S S R D F L G G L A F R V F H C T

15 (Secuencia de nucleótidos - SEQ ID NO: 49; secuencia de proteínas - SEQ ID NO: 50).

Síndrome de Bernard-Soulier (BSS) - cisteína a arginina en el resto 24 en la glucoproteína IX de la membrana plaquetaria (mutación T>C):

atgcctgcctggggagccctgttctctgctctgggccacagcagaggccaccaaggactgc
 M P A W G A L F L L W A T A E A T K D C
 cccagcccccgtacctgccgcgccctggaaaccatggggctgtgggtggactgcaggggc
 P S P R T C R A L E T M G L W V D C R G
 cacggactcacggccctgctgcccctgcccggcccgcaccgccaccttctgctggccaac
 H G L T A L P A L P A R T R H L L L A N

(Secuencia de nucleótidos - SEQ ID NO: 51; secuencia de proteínas - SEQ ID NO: 52).

20 Hiperqueratosis epidermolítica (EHK) - mutación de leucina a prolina en el resto 161 en queratina 1 (mutación T>C):

ggttatggctcctgtctgccctcctgggtggcatacaagaagtcactatcaaccagagccct
 G Y G P V C P P G G I Q E V T I N Q S P
 cttcagccccctcaatgtggagattgacctgagatccaaaagggtgaagtctcgagaaaagg
 L Q P L N V E I D P E I Q K V K S R E R

(Secuencia de nucleótidos - SEQ ID NO: 53; secuencia de proteínas - SEQ ID NO: 54).

Enfermedad pulmonar obstructiva crónica (EPOC) - mutación de leucina a prolina en el resto 54 en α_1 -antitripsina (mutación T>C):

gtctccctggctgaggatccccagggagatgctgcccagaagacagatacatcccaccat
 V S L A E D P Q G D A A Q K T D T S H H
 gatcaggatcacccaaccttcaacaagatcaccccccaacccggctgagttcgccttcagc
 D Q D H P T F N K I T P N P A E F A F S
 ctataaccgccagctggcacaccagtccaacagcaccaatatcttcttccccagtgagc
 L Y R Q L A H Q S N S T N I F F S P V S

25

ES 2 754 433 T3

(Secuencia de nucleótidos - SEQ ID NO: 55; secuencia de proteínas - SEQ ID NO: 56).

Enfermedad pulmonar obstructiva crónica (EPOC) - mutación de leucina a prolina en el resto 78 en α 1-antiquimiotripsina (mutación T>C):

gctctccgcccaacgtggacttcgctttcagcctgtacaagcagtttagtctctgaaggccct

A S A N V D F A F S L Y K Q L V L K A P
gataagaatgtcatcttctccccaccgagcactctccaccgcttggccttctctgtctctg
D K N V I F S P P S I S T A L A F L S L
ggggcccataataaccacctgacagagattctcaaaggcctcaagttctacctcaaggag
5 G A H N T T L T E I L K G L K F Y L T E

(Secuencia de nucleótidos - SEQ ID NO: 89; secuencia de proteínas - SEQ ID NO: 90).

Neuroblastoma (NB) - mutación de leucina a prolina en el resto 197 en caspasa-9 (mutación T>C):

ggcactgctcattatcaacaatgtgaacttctgcccgtgagtcggggtccgcacccgc
G H C L I I N N V N F C R E S G L R T R
actggctccaacatcgactgtgagaagttgcccgcgtcgttctctctcgccgcatttcatg
T G S N I D C E K L R R R F S S P H F M
gtggaggtgaagggcgacctgactgccaaagaaaatgggtgctggctttgtggagctggcg
V E V K G D L T A K K M V L A L L E L A

(Secuencia de nucleótidos - SEQ ID NO: 57; secuencia de proteínas - SEQ ID NO: 58).

10 Enfermedad de Charcot-Marie-Tooth tipo 4J - mutación de isoleucina a treonina en el resto 41 en FIG4 (mutación T>C):

actagagctagatactttctagttgggagcaataatgcagaaacgaaatatcgtgtcttg
T R A R Y F L V G S N N A E T K Y R V L
aagactgatagaaacagaaccaaagatttgggtcataattgatgacaggcatgtctatact
K T D R T E P K D L V I I D D R H V Y T
caacaagaagtaaggggaacttcttggccgcttggatcttggaaatagaacaaagatggga
Q Q E V R E L L G R L D L G N R T K M G

(Secuencia de nucleótidos - SEQ ID NO: 59; secuencia de proteínas - SEQ ID NO: 60).

15 Enfermedad de von Willebrand (vWD) - mutación de cisteína a arginina en el resto 1272 en factor de von Willebrand (mutación T>C):

acagatgccccgggtgagccccaccactctgtatgtggaggacatctcggaaaccgcccgttg
T D A P V S P T T L Y V E D I S E P P L
cacgatttctaccgcagcaggctactggacctgggtcttctctgctggatggctcctccagg
H D F Y R S R L L D L V F L L D G S S R
ctgtccgaggtgagtttgaagtgctgaaggcctttgtggtggacatgatggagcggctg
L S E A E F E V L K A F V V D M M E R L

(Secuencia de nucleótidos - SEQ ID NO: 61; secuencia de proteínas - SEQ ID NO: 62).

Miotonía congénita - mutación de cisteína a arginina en la posición 277 en el gel del canal de cloruro muscular CLCN1 (mutación T>C):

atctgtgctgctgtcctcagcaaattcatgtctgtgttctgccccgtatatgagcagcca
I C A A V L S K F M S V F C G V Y E Q P
tactactactctgatcctgacgggtgggctgtgctgtgggagtcggccggttgttttggg
Y Y Y S D I L T V G C A V G V G R C F G
20 acaccacttggaggagtgtctatcttagcatcgagggtcacctccacctacttttgtgttccg
T P L G G V L F S I E V T S T Y F A V R

ES 2 754 433 T3

(Secuencia de nucleótidos - SEQ ID NO: 63; secuencia de proteínas - SEQ ID NO: 64).

Amiloidosis renal hereditaria - mutación de codón de terminación a arginina en el resto 111 en apolipoproteína All (mutación T>C):

```
tactttgaaaagtcaaaggagcagctgacacccctgatcaagaaggctggaacggaactg
Y F E K S K E Q L T P L I K K A G T E L
gttaacttcttgagctatcttctggaacttggaaacacagcctgccaccagcgaagtgtc
V N F L S Y F V E L G T Q P A T Q R S V
cagcaccattgtcttccaaccccagctggcctctagaaccccactggccagtcctagag
Q H H C L P T P A G L - N T H W P V L E
```

5 (Secuencia de nucleótidos - SEQ ID NO: 65; secuencia de proteínas - SEQ ID NO: 66).

Cardiomiopatía dilatada (DCM) - mutación de triptófano a arginina en la posición 148 en el gen FOXD4 (mutación T>C):

```
ccgcacaagcgcctcacgctcagcggcatctgcgccttcattagtgaccgcttcccctac
P H K R L T L S G I C A F I S D R F P Y
taccgccgcaagttccccggcggcagaacagcatccgccacaacctctcgcgtgaacgac
Y R R K F P A R Q N S I R H N L S L N D
tgcttcgtcaagatcccccgagccggggccgcccaggcaagggcaactactggagcctg
C F V K I P R E P G R P G K G N Y W S L
```

(Secuencia de nucleótidos - SEQ ID NO: 67; secuencia de proteínas - SEQ ID NO: 68).

10 Linfedema hereditario - mutación de histidina a arginina en el resto 1035 en tirosina cinasa de VEGFR3 (mutación A>G):

```
gctgaggacctgtggctgagcccgcctgaccatggaagatcttgtctgctacagcttccag
A E D L W L S P L T M E D L V C Y S F Q
gtggccagagggatggagttcctggcttccccgaaagtgcacccgagagacctggctgct
V A R G M E F L A S R K C I R R D L A A
cggaacattctgctgtcggaaagcagcgtgggtaagatctgtgactttggccttggcccgg
R N I L L S E S D V V K I C D F G L A R
```

(Secuencia de nucleótidos - SEQ ID NO: 69; secuencia de proteínas - SEQ ID NO: 70).

Enfermedad de Alzheimer familiar - mutación de isoleucina a valina en el resto 143 en presenilina 1 (mutación A>G):

```
gataccgagactgtgggcccagagagccctgcactcaattctgaatgctgccatcatgatc
D T E T V G Q R A L H S I L N A A I M I
agtgtcgttgttgcacgactatcctcctgggtggttctgtataaatacaggtgctataag
S V V V V M T I L L V V L Y K Y R C Y K
gtcatccatgcctggccttattatcatctctattgttgcgttctttttttcattcatt
V I H A W L I I S S L L L L F F F S F I
```

15

(Secuencia de nucleótidos - SEQ ID NO: 71; secuencia de proteínas - SEQ ID NO: 72).

Enfermedad priónica - mutación de metionina a valina en el resto 129 en proteína priónica (mutación A>G):

```
aagccgagtaagccaaaaaccaacatgaagcacatggctgggtgctgcagcagctggggca
K P S K P K T N M K H M A G A A A A G A
gtgggtggggggccttggcggctacgtgctgggaagtggccatgagcaggcccatcatacat
V V G G L G G Y V L G S A M S R P I I H
ttcggcagtgactatgaggaccgttactatcgtgaaaacatgcaccggttaccccaaccaa
F G S D Y E D R Y Y R E N M H R Y P N Q
```

(Secuencia de nucleótidos - SEQ ID NO: 73; secuencia de proteínas - SEQ ID NO: 74).

Síndrome crónico, infantil, neurológico, cutáneo, articular (CINCA) - mutación de tirosina a cisteína en el resto 570 en criopirina (mutación A>G):

```
cttcccagccgagacgtgacagtcctttctggaaaactatggcaaattcgaaaaggggtgt
L P S R D V T V L L E N Y G K F E K G C
ttgatttttgttgtagctttcctctttggcctggtaaacaggagaggacctcctacttg
L I F V V R F L F G L V N Q E R T S Y L
```

(Secuencia de nucleótidos - SEQ ID NO: 75; secuencia de proteínas - SEQ ID NO: 76).

5 Miopatía relacionada con desmina (DRM) - mutación de arginina a glicina en el resto 120 en αB cristalina (mutación A>G):

```
gtgaagcacttctccccagaggaactcaaagttaagggtgttgggagatgtgattgaggtg
V K H F S P E E L K V K V L G D V I E V
catggaaaacatgaagagcgccaggatgaacatggtttcatctccagggagttccacggg
H G K H E E R Q D E H G F I S R E F H G
aaataccggatcccagctgatgtagacctctcaccattacttcatccctgtcatctgat
K Y R I P A D V D P L T I T S S L S S D
```

(Secuencia de nucleótidos - SEQ ID NO: 77; secuencia de proteínas - SEQ ID NO: 78).

Beta-talasemia - un ejemplo es la mutación de leucina a prolina en el resto 115 en hemoglobina B.

```
gagctgcaactgtgacaagctgcacgtggatcctgagaacttcaggctcctgggcaacgtg
E L H C D K L H V D P E N F R L L G N V
ctggctctgtgtgcgggccatcactttggcaagaattcaccaccagtgaggtgcc
L V C V P A H H F G K E F T P P V Q A A
tatcagaaagtgggtggctgggtgtggctaatgccctggcccacaagtatcactaagctcgc
Y Q K V V A G V A N A L A H K Y H - A R
```

10

(Secuencia de nucleótidos - SEQ ID NO: 79; secuencia de proteínas - SEQ ID NO: 80).

Se debe entender que las secuencias proporcionadas anteriormente son a modo de ejemplo y no pretende ser limitantes del alcance de la presente divulgación. Las secuencias adecuadas adicionales de mutaciones puntuales que están asociadas con enfermedad y susceptibles a corrección por proteína de fusión de Cas9:enzima/dominio de edición de ácidos nucleicos, así como secuencias guía de ARN adecuadas, serán evidentes para los expertos en la técnica basándose en la presente divulgación.

15

Sistemas indicadores

Algunos aspectos de la presente divulgación proporcionan un sistema indicador que se puede usar para detectar actividad de desaminasa de las proteínas de fusión descritas en el presente documento. El sistema indicador puede ser un ensayo basado en luciferasa en el que la actividad de desaminasa conduce a la expresión de luciferasa. Para minimizar el impacto de la posible promiscuidad de sustrato del dominio de desaminasa (por ejemplo, el dominio de ayuda), se minimiza el número de restos que podrían ser involuntariamente dirigidos para la desaminación (por ejemplo, restos C inespecíficos que podrían residir posiblemente en el ADNmc dentro del sistema indicador). Un resto diana previsto se puede localizar en un codón de iniciación mutado de ACG del gen luciferasa que es incapaz de iniciar la traducción. La actividad de desaminasa deseada da como resultado una modificación ACG>AUG, permitiendo así la traducción de luciferasa y la detección y cuantificación de la actividad de desaminasa.

20

25

Para minimizar los restos de C monocatenarios, se puede insertar una secuencia conductora entre el codón de iniciación mutado y el inicio del gen luciferasa que consiste en una extensión de restos Lys (AAA), Asn (AAT), Leu (TTA), Ile (ATT, ATA), Tyr (TAT) o Phe (TTT). Los mutantes resultantes se pueden probar para garantizar que la secuencia conductora no afecte adversamente la expresión o actividad de luciferasa. También se puede determinar la actividad de luciferasa de fondo con el codón de iniciación mutado.

30

Se puede usar el sistema indicador para probar muchos ARNgu diferentes, por ejemplo, para determinar qué resto(s) con respecto a la secuencia de ADN diana dirigirán la desaminasa respectiva (por ejemplo, enzima AID) (Figura 3). Debido a que no se conoce el tamaño de la burbuja de Cas9-ADN, también se pueden probar ARNgu que dirigen la cadena no de molde para evaluar efectos inespecíficos de una proteína de fusión de Cas9-desaminasa específica. Dichos ARNgu se pueden diseñar de forma que el codón de iniciación mutado no aparezca las bases con el ARNgu.

35

Una vez se han identificado las proteínas de fusión que son capaces de modificaciones de C a U específicas de sitio programables, sus actividades se pueden caracterizar adicionalmente. Los datos de los ensayos de luciferasa se pueden integrar, por ejemplo, en mapas de calor que describen qué nucleótidos, con respecto al ADN diana de ARN_g, están siendo dirigidos para la desaminación por una proteína de fusión específica. La posición que da como resultado la mayor actividad en el ensayo de luciferasa para cada fusión se puede considerar la posición "diana", mientras que todas las otras se consideran posiciones inespecíficas.

En algunas realizaciones, se proporcionan fusiones de Cas9 con diversas enzimas APOBEC3, o dominios de desaminasa de las mismas. En algunas realizaciones, se proporcionan proteínas de fusión de Cas9 con otras enzimas de edición de ácidos nucleicos o dominios catalíticos, que incluyen, por ejemplo, enzimas de edición de ARN_{mc}, tales como las citidina desaminasas APOBEC1 y ACF1/ASF, así como la familia ADAT de adenosina desaminasas,³⁸ que se puede usar para la actividad de edición de ADN_{mc} cuando se fusiona con Cas9. La actividad de dichas proteínas de fusión se puede probar usando los mismos sistemas indicadores y ensayos descritos anteriormente.

Se proporciona un sistema indicador en el presente documento que incluye un gen indicador que comprende un codón de iniciación desactivado, por ejemplo, una mutación en la cadena de molde desde 3'-TAC-5' hasta 3'-CAC-5'. Tras la desaminación satisfactoria de C diana, el ARNm correspondiente se transcribirá como 5'-AUG-3' en lugar de 5'-GUG-3', que permite la traducción del gen indicador. Los genes indicadores adecuados serán evidentes para los expertos en la técnica.

La descripción de realizaciones a modo de ejemplo de los sistemas indicadores anteriores se proporciona para fines de ilustración solo y no pretende ser limitante. También están englobados por la presente divulgación sistemas indicadores adicionales, por ejemplo, variaciones de los sistemas descritos a modo de ejemplo en detalle anteriormente.

Ejemplos

Ejemplo 1: Proteínas de fusión

Se proporcionan a continuación proteínas de fusión de Cas9:desaminasa a modo de ejemplo:

Fusión de Cas9:AID humana (extremo C)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFK
 VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
 SFFHRLEESFLVEEDDKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLA
 LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDKAIL SARLSKS
 RRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQ
 IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
 PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTR
 KSEETITPWNFEVVDK GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
 VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
 GTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKR
 RRYTGWGRLSRKLINGIRDKQSGKITLDLFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSG
 QGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNS
 RERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVD
 AIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
 KAERGGSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLTKSKLV
 SDFRKFDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAS
 EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
 EKGKSKLKS VKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKKHYLDEIEQI
 SEFSKRVILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRRK

YTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD**GGGGSGGGSGGGGS**YVVKRRDSATSFSL
 DFGYLRNKNKGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSL
 RIFTARLYFCEDRKAEP EGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAW EGLHEN
 SVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 30)

(subrayado: señal de localización nuclear; subrayado doble: señal de exportación nuclear, negrita: secuencia conectora)

Fusión de Cas9:AID humana (extremo N)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELLFL
 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEP E
 GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAW EGLHENSVRLSRQLRRILLP**GGGG**
SGGGSGGGGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHS IKKNLIGALL
 FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
 DNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLF
 GNLIALSGLTPNFKSNFDLAEDAKLQLSKDYYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGY
 IDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
 RQEDFY PFLKDNREKIEKILTFRIPIYVGLARGNSRFAMTRKSEETITPWNFEVVDKGA
 SAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNE
 ENEDILEDIVLTLTLFEDREMIEERLKYAHLFDKVMKQLKRRRYTGWGRLSRKLINGIRD
 KQSGKTI LDFLKS DGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IK
 KGILQTVKVVDELVKVMGRHKPENIVIAMARENQTQKGQKNSRERMKRIEEGIKELGSQIL
 KEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLT
 RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRO
 LVETROITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYK VREINNYH
 HAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMI AKSEQEIGKATAKYFFYSNIMN
 FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFS
 KESILPKRNSDKLIARKKDWDPKYG GFDSPVAYSVLVVAKVEKGKSKKLKSVKELLGITI
 MERSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
 KYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL S
 AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI T
 LYETRIDLSQLGGD (SEQ ID NO: 31)

5

(subrayado: señal de localización nuclear; negrita: secuencia conectora)

Fusión de Cas9:AID de ratón (extremo C)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFK
 VLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
 SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
 LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSK
 SRLENLIAQLPGEKKNGLFGNLIALSGLTPNFKSNFDLAEDAKLQLSKDYYDDDLNLLAQ
 IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
 PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTR
 KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEYFTVYNELTKVKY
 VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
 GTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKR
 RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSG
 QGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
 RERMKRIIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
 AIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
 KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV
 SDFRKDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIAS
 EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAVK
 EKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFYLAHYEKLKGPSPEDNEQKQLFVEQHKHYLDEIEQI
 SEFSKRVIADANLDKVL SAYNKHRDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKR
 YTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD**GGGGSGGGSGGGGS**YVVKRRDSATSCSL
 DFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSL
 RIFTARLYFCEDRKAEPGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHEN
 SVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 32)

(subrayado: señal de localización nuclear; negrita: secuencia conectora; subrayado doble: señal de exportación nuclear)

Fusión de Cas9:APOBEC-3G humana (extremo N)

SPKKKRKVEASMEKLYHPMRFFHWFSSKWRKLHRDQEYEV TWYISWSPCTKCTRDMATFLAE
 DPKVTLTIFVARLYYFWDPDYQEA LRSLCQKRDRGPRATMKIMNYDEFQHCWSKFVYSQRELF
 EPWNNLPKYIILLHIMLGEILRHSMDPPTFTFNENNEPWVRGRHETYLCEVERMHNDTWVL
 LNQRGFLCNQAPHKHGFLEGRHAELCFLDVI PFWKLDLDQDYRVTCFTSWSPCFSCAQEMA
 KFISKNKHVSLCIFTARIYDDQGRCEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPF
 QPWDGLDEHSQDLSGRRLRAILQNQEN**SPKKKRKVEASSPKKKRVEAS**KKKYSIGLAIGTNSV
 GWAVITDEYKVP SKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRKNR
 ICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
 KLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPIN
 ASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSGLTPNFKSNFDLAEDAKL
 QLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
 EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
 LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPI
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHK
 LLEYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
 FDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIH
 DD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIV
 IEMARENQTTQKGQKNSRERMKRIIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
 YVDQELDINRLSDYDVD AIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWR
 QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN

5

DKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTTALIKKYPKLESEF
 VYGDYKVDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
 EIVWDKGRDFATVRKVLSPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSPTVAYSVLVVAKEKSKKLSVKELLGITIMERSSEKPNPIDFLEAKGYKEVKKDL
 IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQ
 LFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNL
 GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD (SEQ ID NO: 33)

(subrayado: señal de localización nuclear; negrita: conector (1 NLS),

Fusión de Cas9:APOBEC-1 humana (extremo N)

SPKKKRKVEASMTSEKGPSTGDPTLRRRIEPWEFDVFDPRELRKEACLLYEIKWGMRSRKIW
 RSSGKNTTNHVEVNF IKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLV
 IYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWM
 MLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCYQTI PPHILLATGLIHPSVAWRS
PKKKRKEASSPKKKRKEASDKKYSIGLAI GTSVGVAVITDEYKVP SKKFKVLGNTDRHS
 IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEES
 FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
 HFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQ
 LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAQLQSKD TYDDDDLNL LAQIGDQYADLF
 LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRQRTFDNGSIPHQI
 HLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYVVGPLARGNSRFAMWTRKSEETITPW
 NFEEVVDK GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
 FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRL
 SRKLINGIRDKQSGKTI LDFLKS DGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
 ANLAGSPA IKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
 GIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLK
 DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLE
 LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF
 YKREINNYHHAHDAYLNAVVGTTALIKKYPKLESEFVYGDYKVDVVRKMIKSEQEIGKATA
 KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKK
 KTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKSKKLS
 VKELLGITIMERSSEKPNPIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKRMLASAGEL
 QKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVIL
 ADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD

5 (subrayado: señal de localización nuclear; negrita: conector (1 NLS), (SEQ ID NO: 92)

Fusión de Cas9:ADAT1 humana (extremo N)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCSMGTGTKCIGQSKMRKNGDILNDSHAEVIARR
SFQRYLLHQQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCGDASII PMLEFEDQ

PCCPVFRNWAHNSSEASSNLEAPGNERKCEDPDSVPTKKMRLEPGTAAREVTNGAAHHQSFGKQKSGPISPGIHSCLTVEGLATVTRIAAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFHQVGLLRVKPGRGDRTRSMSCSDKMARWNVVLCGCGALLMHLLEEPIYLSAVVIGKCPYSQEAMQRALIGRCQNVSALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDKWPFLSLRVQKLDYQEYKEAASSYQEAWSLTKRQVFGSWIRNPPDYHQF**GGGGSGGGSGGGG**DKKYSIGLAIGTNSVGWAVITDEYKVPSPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSHEHIANLAGSPAIKKGIQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIQQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIINNYHHAHDAYLNAVVGTAALIKKYPKLESEFVYGDYKVVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VVKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKLLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKKHYLDEIEQISEFSKRVI LADANL DKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD (SEQ ID NO: 35)

(subrayado: señal de localización nuclear; negrita: secuencia conectora)

Fusión de Cas9:ADAT1 humana (extremo)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPSPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKR

RRYTGWGRLSRKLINGIRDKQSGKTIIDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSG
 QGDSLHEHIANLAGSPAIKKGILOTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
 RERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVD
 AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
 KAERGGSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIITLKSCLV
 SDFRKFDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIAKS
 EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 SMPQVNIVKKEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKV
 EKGKSKLKSVELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIKLPHYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEEQI
 SEFSKRVIADANLDKVL SAYNKHRDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKR
 YTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD**GGGGSGGGGS**SMGTGTKCIGQSKMRKNGD
 ILNDSHAEVIARRSFQRYLLHQQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCG
 DASIIPMLEFEDQPCPVFRNWAHNSSEASSNEAPGNERKCEDPDSPTKMRLEPGTAA
 REVINGAAHHQSFQKQSGPISPGIHSCLTVEGLATVTRIPGSAKVIDVYRTGAKCVPGE
 AGDSGKPGAAHFQVGLLRVKPGRGDRTRSMSCSDKMARWNVLCQGALLMHLLLEPIYLSAV
 VIGKCPYSQEQALIGRCQNVSALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRL
 VPCGAAISWSAVPEQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDK
 WPHSLRVQKLDTYQEYKEAASSYQEAWSTLRKQVFGSWIRNPPDYHQF (SEQ ID NO:
 36)

(subrayado: señal de localización nuclear; negrita: secuencia conectora)

Ejemplo 2: Corrección de una mutación puntual PI3K por una proteína de fusión de Cas9

5 Se corrige una mutación puntual A3140G en el exón 20 del gen PI3KCA, que da como resultado una sustitución de aminoácidos H1047R en la proteína PI3K, poniendo en contacto un ácido nucleico que codifica la proteína mutante con una proteína de fusión de Cas9:AID (SEQ ID NO: 30) o Cas9:APOBEC1 (SEQ ID NO: 92) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en el gen PI3KCA codificante. La mutación puntual A3140G se confirma por PCR genómica de la secuencia respectiva del exón 20, por ejemplo, generación de un amplicón por PCR de los nucleótidos 3000-3250, y posterior secuenciación del amplicón PCT.

10 Se ponen en contacto células que expresan una proteína PI3K mutante que comprende una mutación puntual A3140G en el exón 20 con una construcción de expresión que codifica la proteína de fusión Cas9:AID (SEQ ID NO: 30) o Cas9:APOBEC1 (SEQ ID NO: 92) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en la hebra no codificante de gen PI3KCA codificante. El ARN_g es de la secuencia 5'-
 15 aucggauctauuuugacucguuuuagagcuagaaaua
 gcaaguuaaaauaaaggcuagucgguuaucacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 81); 5'-
 ucggaucuuuuugacucgguuuuagagcuagaaauagcaaguuaaaauaaaggcuagucgguuaucacuugaaaa
 guggcaccgagucggugcuuuuu-3' (SEQ ID NO: 82); 5'-cuuagauaaaacugagcaagguuuuagagcuagaa
 auagcaaguuaaaauaaaggcuagucgguuaucacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 83); 5'-
 20 aucauuuuugacucguucucgguuuuagagcuagaaauagcaaguuaaaauaaaggcuagucgguuaucacuug
 aaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 84); 5'-uaaaacugagcaagaggcuuuuuuagagcu
 gaaaagcaaguuaaaauaaaggcuagucgguuaucacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 85); 5'-
 ugguggcuggacaacaaaaaguuuuuagagcuagaaauagcaaguuaaaauaaaggcuagucgguuaucac
 cuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 86); 5'-gcuggacaacaaaaauggauguuuu
 gagcuagaaauagcaaguuaaaauaaaggcuagucgguuaucacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 87); o
 25 5'-guguuaauuugucguacguuuuagagcuagaaauagcaaguuaaaauaaaggcu
 guccguuaucacuugaaaaaguggcaccgagucggugcuuuuu (SEQ ID NO: 88).

La actividad de citosina desaminasa de la proteína de fusión Cas9:AID o Cas9:APOBEC1 da como resultado la desaminación de la citosina que está apareada con base con G3140 mutante a uridina. Después de una ronda de replicación, se restaura A3140 no mutante. Se extrae el ADN genómico de las células tratadas y se amplifica un amplicón de PCR de nucleótidos 3000-3250 con cebadores de PCR adecuados. La corrección de la mutación puntual A3140G después del tratamiento de las células con la proteína de fusión se confirma por secuenciación del amplicón de PCR.

Ejemplo 3: Corrección de una mutación puntual de presenilina 1 por una proteína de fusión de Cas9

35 Se corrige una mutación puntual A->G en el codón 143 del gen presenilina 1 (PSEN1), dando como resultado una sustitución de aminoácidos I143V en la proteína PSEN1 poniendo en contacto un ácido nucleico que codifica la

proteína PSEN1 mutante con una proteína de fusión Cas9:AID (SEQ ID NO: 30) o Cas9:APOBEC1 (SEQ ID NO: 92) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en el gen PSEN1 codificante. Véase, por ejemplo, Gallo et. al., J. Alzheimer's disease. 2011; 25: 425-431 para una descripción de una mutación I143V en PSEN1 a modo de ejemplo asociada a enfermedad de Alzheimer familiar. La mutación puntual A->G se confirma por PCR genómica de la secuencia respectiva de PSEN1, por ejemplo, generación de un amplicón de PCR de aproximadamente 100-250 nucleótidos alrededor del exón 143, y posterior secuenciación del amplicón PCT.

Se ponen en contacto células que expresan la proteína PSEN1 mutante con una construcción de expresión que codifica la proteína de fusión de Cas9:AID (SEQ ID NO: 30) o Cas9:APOBEC1 (SEQ ID NO: 92) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en la cadena no codificante del gen PSEN1 codificante. La actividad de citosina desaminasa de la proteína de fusión de Cas9:AID o Cas9:APOBEC1 da como resultado la desaminación de la citosina que está apareada con base con G mutante en el codón 143 a uridina. Después de una ronda de replicación, se restaura A no mutante. Se extrae el ADN genómico de las células tratadas y se amplifica un amplicón de PCR de 100-250 nucleótidos con cebadores de PCR adecuados. Se confirma la corrección de la mutación puntual A->G después del tratamiento de las células con la proteína de fusión por secuenciación del amplicón de PCR.

Ejemplo 4: Corrección de una mutación de puntual α_1 -antitripsina por una proteína de fusión de Cas9

Se corrige una mutación puntual T->C en el codón 55 del gen α_1 -antitripsina, que da como resultado una sustitución de aminoácidos L55P en la proteína α_1 -antitripsina, poniendo en contacto un ácido nucleico que codifica la proteína α_1 -antitripsina mutante con una proteína de fusión de Cas9:ADAT1 (SEQ ID NO: 35 o 36) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en el gen α_1 -antitripsina codificante. Véase, por ejemplo, Poller et al., Genomics. 1993; 17: 740-743 para una descripción más detallada de una mutación T->C del codón 55 a modo de ejemplo asociada a enfermedad pulmonar obstructiva crónica (EPOC). La mutación puntual T->C se confirma por PCR genómica de la secuencia de α_1 -antitripsina respectiva que codifica el codón 55, por ejemplo, generación de un amplicón de PCR de aproximadamente 100-250 nucleótidos, y posterior secuenciación del amplicón de PCT.

Se ponen en contacto células que expresan la proteína α_1 -antitripsina mutante con una construcción de expresión que codifica la proteína de fusión de Cas9:AID (SEQ ID NO: 30) o Cas9:APOBEC1 (SEQ ID NO: 92) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al nucleótido mutado en el codón 55 en la cadena codificante en el gen α_1 -antitripsina codificante. La actividad de citosina desaminasa de la proteína de fusión Cas9:ADAT1 da como resultado la desaminación de la citosina mutante a uridina corrigiendo así la mutación. Se extrae el ADN genómico de las células tratadas y se amplifica un amplicón de PCR de 100-250 nucleótidos con cebadores de PCR adecuados. La corrección de la mutación puntual T->C en el codón 55 del gen α_1 -antitripsina después del tratamiento de las células con la proteína de fusión se confirma por secuenciación del amplicón de PCR.

Ejemplo 5: Corrección de una mutación puntual del factor de von Willebrand por una proteína de fusión de Cas9

Se corrige una mutación puntual T->C en el codón 509 del gen factor de von Willebrand, que da como resultado una sustitución de aminoácidos C509A en la proteína factor de von Willebrand, poniendo en contacto un ácido nucleico que codifica la proteína factor de von Willebrand mutante con una proteína de fusión de Cas9:ADAT1 (SEQ ID NO: 35 o 36) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en la cadena codificante del gen factor de von Willebrand codificante. Véase, por ejemplo, Lavergne et al., Br. J. Haematol. 1992; 82: 66-7, para una descripción de una mutación C509A de factor de von Willebrand a modo de ejemplo asociada a enfermedad de von Willebrand (vWD). La mutación puntual T->C se confirma por PCR genómica de la secuencia genómica respectiva del gen factor de von Willebrand, por ejemplo, generación de un amplicón de PCR de aproximadamente 100-250 nucleótidos alrededor del exón 509, y posterior secuenciación del amplicón PCT.

Se ponen en contacto células que expresan la proteína factor de von Willebrand mutante con una construcción de expresión que codifica la proteína de fusión de Cas9:ADAT1 (SEQ ID NO: 35 o 36) y un ARN_g apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en la cadena codificante del gen factor de von Willebrand codificante. La actividad de citosina desaminasa de la proteína de fusión de Cas9:ADAT1 da como resultado la desaminación de la citosina mutante en el codón 509 a uridina, corrigiendo así la mutación. Se extrae el ADN genómico de las células tratadas y se amplifica un amplicón de PCR de 100-250 nucleótidos con cebadores de PCR adecuados. La corrección de la mutación puntual T->C en el codón 509 del gen factor de von Willebrand después del tratamiento de las células con la proteína de fusión se confirma por secuenciación del amplicón de PCR.

Ejemplo 6: Corrección de una mutación puntual en caspasa 9 por una proteína de fusión de Cas9 - neuroblastoma

Se corrige una mutación puntual T->C en el codón 197 del gen caspasa-9, que da como resultado una sustitución de aminoácidos L197P en la proteína caspasa-9, poniendo en contacto un ácido nucleico que codifica la proteína caspasa-9 mutante con una proteína de fusión de Cas9:ADAT1 (SEQ ID NO: 35 o 36) y un ARN_g apropiadamente

diseñado que dirige la proteína de fusión al sitio de mutación en la cadena codificante del gen de caspasa-9 codificante. Véase, por ejemplo, Lenk et al., PLoS Genetics. 2011; 7: e1002104, para una descripción de una mutación L197P de caspasa-9 a modo de ejemplo asociada a neuroblastoma (NB). La mutación puntual T->C se confirma por PCR genómica de la secuencia genómica de caspasa-9 respectiva, por ejemplo, generación de un amplicón de PCR de aproximadamente 100-250 nucleótidos alrededor del exón 197, y posterior secuenciación del amplicón PCT.

5 Se ponen en contacto células que expresan la proteína caspasa-9 mutante con una construcción de expresión que codifica la proteína de fusión de Cas9:ADAT1 (SEQ ID NO: 35 o 36) y un ARNgü apropiadamente diseñado que dirige la proteína de fusión al sitio de mutación en la cadena codificante del gen caspasa-9 codificante. La actividad de citosina desaminasa de la proteína de fusión de Cas9:ADAT1 da como resultado la desaminación de la citosina mutante en el codón 197 a uridina, corrigiendo así la mutación. Se extrae el ADN genómico de las células tratadas y se amplifica un amplicón de PCR de 100-250 nucleótidos con cebadores de PCR adecuados. La corrección de la mutación puntual T->C en el codón 197 del gen caspasa-9 después del tratamiento de las células con la proteína de fusión se confirma por secuenciación del amplicón de PCR.

15 **Ejemplo 7: Actividad de desaminasa de dos proteínas de fusión de dCas9-APOBEC1**

Se generaron dos proteínas de fusión de dCas9-APOBEC1 con diferentes conectores:

rAPOBEC1_GGS_dCas9:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINWGRHSIWRHTSQNTNKHV
EVNFIEKFTTERTYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHAD
PRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCII
LGLPPCLNILRRKQPQLTFFTTIALQSCHYQRLPPHILWATGLKGGSMDDKKYSIGLAIGTNSV
GWAVITDEYKVPKFKVGLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
ICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALSGLTLPNFKSNFDLAEDAKL
QLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHDQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFKLDNREKIEKILTFRIPIY
YVGPLARGNSRFAMTRKSEETITPWNFEVVDKGSASQSFIERMTNFDKLNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIIDFLKSDGFANRNFMLIHL
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILOTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTTRSDKNRGSNDVNPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLELKDAGFIKQRLVETROITKHVAQILD SRMNTKYDEN
DKLIREVKVITLKSCLVSDFRKDFQFYKVIINNYHHAHDAYLNAVVGTAIHKYPKLESEF
VYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKEKGSKLLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQ
LFEVQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQISITGLYETRIDLSQLGGD (SEQ ID NO:
 94);

subrayado = rAPOBEC1; subrayado doble = dCas9.

rAPOBEC1_(GGG)₃_dCas9:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHV
EVNFIEKFTTTERYFCPNTRCSIWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHAD
PRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCII
LGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSGGSGGSMDDKYSIGLA
IGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDS TDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF
EENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLNLLAQIGDOYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKIL
TFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASAOFSIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLKLIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDFLKSDFANRN
FMQLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTOKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLONEKLYLYYL
QNGRDMYVDQELDINRLSDYDVAIVPOSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKQRLVETROITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYP
KLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVAKVEK GKSKLKS VKELLGITIMERSSEKNPIDFLEAKGYK
EVKCDLIIKLPKYSLFELENGRKRMLASAGELQGNELALPSKYVNFLYLASHYEKLGSP
DNEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

(SEQ ID NO: 95); subrayado = rAPOBEC1; subrayado doble = dCas9.

5 Se examinó la actividad de desaminasa de ambas proteínas de fusión. Se adaptó un ensayo de desaminasa de Nuc. Acids Res. 2014, 42, p. 1095; J. Biol. Chem. 2004, 279, p 53379; J. Virology 2014, 88, p. 3850; y J. Virology 2006, 80, p. 5992.

10 Se insertaron construcciones de expresión que codifican las proteínas de fusión en un plásmido de esqueleto de CMV (plásmido de Addgene 52970; véase Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82). Se expresaron las proteínas de fusión usando un sistema de transcripción/traducción acoplada TNT Quick (Promega). Después de 90 min, se incubaron 5 µL de lisado con sustrato de ADNmc marcado en 5' (Cy3-ATTATTATTATCCGCGGATTTATT TATTTATTTATTTATTT, SEQ ID NO: 96) y UDG (uracil-ADN glicosilasa) a 37 °C durante 3 h. Entonces se añadió una disolución 1 M de NaOH (10 µL) para escindir el ADN en el sitio abásico. Véase la Figura 4. Se resolvió el ADN en un gel al 10 % de TBE-PAGE (Figura 5). También se incluyeron un control negativo, donde pUC19 se incubó en el sistema TNT, y un control positivo, donde el ADN se ha sintetizado con un "U" en lugar de C diana. La Figura 5 ilustra que ambas proteínas de fusión presentan actividad de citosina desaminasa.

REFERENCIAS

- 20 1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. Crit Rev Biochem Mol.2012; 47(3):264-81. PMID: 22530743.
2. Perez-Pinera P, Ousterout DG, Gersbach CA. Advances in targeted genome editing. Curr Opin Chem Biol. 2012; 16(3-4):268-77. PMID: 22819644.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010; 11(9):636-46. PMID: 20717154.
- 25 4. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol. 2013; 14(1):49-55. PMID: 23169466.

5. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. *Nature*. 2013; 495, (7439):50-1. PMID: 23467164.
6. Pan Y, Xia L, Li AS, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol*. 2013; 55(1):54-62. PMID: 23089945.
- 5 7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods*. 2012; 9(1):27. PMID: 22312638.
8. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA*. 2008; 105(15):5809-14. PMID: 18359850.
9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*. 1999; 22(3):231-8. PMID: 10391209.
- 10 10. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002; 43(6):1565-75. PMID: 11952905.
11. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods*. 2013; 10(10):957-63. PMID: 24076990.
- 15 12. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Shou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*. 2011; 18(5):529-36. PMID: 21460843.
- 20 13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010; 327(5962):167-70. PMID: 20056882.
14. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012; 482(7385):331-8. PMID: 22337052.
- 25 15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol*. 2013; 21(11):562-7. PMID: 24095303.
16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152(5):1173-83. PMID: 23452860.
17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013; 10(10):973-6. PMID: 23892895.
- 30 18. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013; 31(9):833-8. PMID: 23907171.
19. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2):442-51. PMID: 23849981.
- 35 20. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013; 8(11):2180-96. PMID: 24136345.
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339(6121):823-6. PMID: 23287722.
- 40 22. Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science*. 1996; 273(5280):1386-9. PMID: 8703073.
23. Tagalakis AD, Owen JS, Simons JP. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev*. 2005; 71(2):140-4. PMID: 15791601.
- 45 24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci*. 2002; 7(10):435-40. PMID 12399177.
25. Britt AB, May GD. Re-engineering plant gene targeting. *Trends Plant Sci*. 2003; 8(2):90-5. PMID: 12597876.

26. Vagner V, Ehrlich SD. Efficiency of homologous DNA recombination varies along the *Bacillus subtilis* chromosome. *J Bacteriol.* 1988; 170(9):3978-82. PMID: 3137211.
27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* 2004; 32(12):3683-8. PMID: 15252152.
28. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol.* 2007; 25(11):1298-306. PMID: 17965707.
29. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 2008; 9(6):229. PMID: 18598372.
30. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol.* 2003; 4(7):631-8.
31. Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* 2004; 3(1):85-9. PMID: 14697763.
32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* 2006; 83(3):195-200. PMID: 16720547.
33. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* 2008; 456(7218):121-4. PMID: 18849968.
34. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* 2009; 284(41). 27761-5. PMID: 19684020.
35. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* 2005; 44(8):2703-15. PMID 15723516.
36. Barbas CF, Kim DH. Cytidine deaminase fusions and related methods. *PCT Int Appl.* 2010; WO 2010132092 A220101118.
37. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013; 65(10):1357-69. PMID: 23026637.
38. Gerber AP, Keller W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci.* 2001; 26(6):376-84. PMID: 11406411.
39. Yuan L, Kurek I, English J, Keenan R. Laboratory-directed protein evolution. *Microbiol Mol Biol Rev.* 2005; 69(3):373-92. PMID: 16148303.
40. Cobb RE, Sun N, Zhao H. Directed evolution as a powerful synthetic biology tool. *Methods.* 2013; 60(1):81-90. PMID: 22465795.
41. Bershtein S, Tawfik DS. Advances in laboratory evolution of enzymes. *Curr Opin Chem Biol.* 2008; 12(2):151-8. PMID: 18284924.
42. Hida K, Hanes J, Ostermeier M. Directed evolution for drug and nucleic acid delivery. *Adv Drug Deliv Rev.* 2007; 59(15):1562-78. PMID: 17933418.
43. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature.* 2011; 472(7344):499-503. PMID: 21478873.
44. Husimi Y. Selection and evolution of bacteriophages in cellstat. *Adv Biophys.* 1989; 25:1-43. PMID: 2696338.
45. Riechmann L, Holliger P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell.* 1997; 90(2):351-60. PMID: 9244308.
46. Nelson FK, Friedman SM, Smith GP. Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology.* 1981; 108(2):338-50. PMID: 6258292.
47. Rakonjac J, Model P. Roles of pIII in filamentous phage assembly. *J Mol Biol.* 1998; 282(1):25-41.
48. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* 1985; 228(4705):1315-7. PMID: 4001944.

49. Sheridan C. Gene therapy finds its niche. *Nat Biotechnol.* 2011; 29(2):121-8. PMID: 21301435.
50. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene.* 2005; 24(8):1477-80. PMID: 15608678.
- 5 51. Ikediobi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teague J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein JN, Stratton MR, Futreal PA, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther.* 2006; 5(11):2606-12. PMID: 17088437.
- 10

EQUIVALENTES Y ALCANCE

Los expertos en la técnica reconocerán, o serán capaces de determinar usando no más de experimentación rutinaria, muchos equivalentes de las realizaciones descritas en el presente documento. El alcance de la presente divulgación no pretende estar limitado a la descripción anterior, sino que es como se expone en las reivindicaciones adjuntas.

15

Artículos tales como "un", "una", "el" y "la" pueden significar uno o más de uno, a menos que se indique lo contrario o sea evidente de otro modo del contexto. Las reivindicaciones o descripciones que incluyen "o" entre dos o más miembros de un grupo se consideran satisfechas si uno, más de uno, o todos los miembros del grupo están presentes, a menos que se indique lo contrario o sea de otro modo evidente del contexto. La divulgación de un grupo que incluye "o" entre dos o más miembros del grupo proporciona realizaciones en las que exactamente un miembro del grupo está presente, realizaciones en las que más de un miembro del grupo está presente, y realizaciones en las que todos los miembros del grupo están presentes.

20

Donde los elementos se presenten como listas, por ejemplo, en formato de grupos de Markush, se debe entender que también se desvela cada subgrupo posible de los elementos, y que cualquier elemento o subgrupo de elementos se puede retirar del grupo. También se observa que el término "que comprende" pretende ser abierto y permite la inclusión de elementos o etapas adicionales. Se debe entender que, en general, donde se hace referencia a una realización, producto, o método que comprende elementos particulares, características, o etapas, también se proporcionan realizaciones, productos, o métodos que consisten, o consisten esencialmente en, dichos elementos, características, o etapas.

25

30

Donde se dan intervalos, se incluyen puntos extremos. Además, se debe entender que, a menos que se indique lo contrario o sea evidente de otro modo del contexto y/o el entendimiento de un experto habitual en la técnica, valores que se expresan como intervalos pueden asumir cualquier valor específico dentro de los intervalos establecidos en algunas realizaciones, al décimo de la unidad del límite inferior del intervalo, a menos que el contexto dicte claramente de otro modo. Para los fines de brevedad, los valores en cada intervalo no han sido individualmente explicados en detalle en el presente documento, pero se entenderá que cada uno de estos valores se proporciona en el presente documento y se puede reivindicar o renunciar a ellos específicamente. También se debe entender que, a menos que se indique lo contrario o sea de otro modo evidente del contexto y/o el entendimiento de un experto habitual en la técnica, los valores expresados como intervalos pueden asumir cualquier subintervalo dentro del intervalo dado, en donde los puntos extremos del subintervalo se expresan al mismo grado de exactitud que el décimo de la unidad del límite inferior del intervalo.

35

40

REIVINDICACIONES

1. Una proteína de fusión que comprende
 - (i) un dominio de Cas9 de nucleasa inactiva; y
 - (ii) un dominio de desaminasa.
- 5 2. La proteína de fusión de la reivindicación 1, en donde el dominio de desaminasa es;
 - (i) una citidina desaminasa; o
 - (ii) una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC); o
 - (iii) una desaminasa de la familia APOBEC 1; o
 - (iv) una citidina desaminasa inducida por activación (AID); o
 - 10 (v) una ACF1/ASE desaminasa; o
 - (vi) una adenosina desaminasa; o
 - (vii) una desaminasa de la familia ADAT.
3. La proteína de fusión de la reivindicación 1 o la reivindicación 2, en donde el dominio de desaminasa se fusiona con el extremo N del dominio de Cas9, o en donde el dominio de desaminasa se fusiona con el extremo C del dominio de Cas9.
- 15 4. La proteína de fusión de una cualquiera de las reivindicaciones 1-3, en donde el dominio de Cas9 y el dominio de desaminasa se fusionan por un conector, opcionalmente en donde el conector comprende un motivo (GGGS)_n (SEQ ID NO: 91), (G)_n, (EAAAK)_n (SEQ ID NO: 5), (GGS)_n, SGSETPGTSESATPES (SEQ ID NO: 93), o (XP)_n, o una combinación de cualquiera de estos, en donde n es independientemente un número entero entre 1 y 30.
- 20 5. La proteína de fusión de la reivindicación 1 o 2, en donde el dominio de desaminasa es una desaminasa de APOBEC 1.
6. La proteína de fusión de la reivindicación 5, en donde la desaminasa de APOBEC 1 comprende la secuencia de aminoácidos de una cualquiera de SEQ ID NOs: 22-24.
- 25 7. La proteína de fusión de una cualquiera de las reivindicaciones 1-6, en donde el dominio de Cas9 de nucleasa inactiva comprende la secuencia de aminoácidos de SEQ ID NO: 34.
8. La proteína de fusión de una cualquiera de las reivindicaciones 1-7, en donde la proteína de fusión comprende la secuencia de aminoácidos de SEQ ID NO: 92.
9. Una proteína de fusión según cualquiera de las reivindicaciones 1-8 para su uso en medicina.
- 30 10. Un método *in vitro* o *ex vivo* de edición de ADN, comprendiendo el método poner en contacto una molécula de ADN con
 - (a) la proteína de fusión de una de las reivindicaciones 1-8; y
 - (b) un ARN guía único (ARNgu) que dirige la proteína de fusión de (a) a una secuencia de ADN diana de la molécula de ADN; en donde la molécula de ADN se pone en contacto con la proteína de fusión y el ARNgu en una cantidad eficaz y en condiciones adecuadas para la desaminación de una base nucleotídica de la molécula de ADN.
- 35 11. El método de la reivindicación 10, en donde la secuencia de ADN diana comprende una secuencia asociada a una enfermedad o trastorno, y en donde la desaminación de la base nucleotídica da como resultado una secuencia que no se asocia a una enfermedad o trastorno, opcionalmente en donde la secuencia de ADN diana comprende una mutación puntual asociada a una enfermedad o trastorno, y en donde la desaminación corrige la mutación puntual.
- 40 12. El método de la reivindicación 10 o la reivindicación 11, en donde la secuencia de ADN diana comprende una mutación puntual T→C o A→G asociada a una enfermedad o trastorno, y en donde la desaminación de la base C o G mutante da como resultado una secuencia que no se asocia a una enfermedad o trastorno.
- 45 13. El método de la reivindicación 11 o la reivindicación 12, en donde la secuencia asociada a la enfermedad o trastorno codifica una proteína, y en donde la desaminación introduce un codón de terminación en la secuencia asociada a la enfermedad o trastorno, dando como resultado una truncación de la proteína codificada.

14. El método de una cualquiera de las reivindicaciones 11-13, en donde la enfermedad o trastorno es fibrosis quística, fenilcetonuria, hiperqueratosis epidermolítica (EHK), enfermedad de Charcot-Marie-Toot de tipo 4J, neuroblastoma (NB), enfermedad de von Willebrand (vWD), miotonía congénita, amiloidosis renal hereditaria, cardiomiopatía dilatada (DCM), linfedema hereditario, enfermedad de Alzheimer familiar, enfermedad priónica, síndrome crónico, infantil, neurológico, cutáneo, articular (CINCA), miopatía relacionada con desmina (DRM), o una enfermedad neoplásica asociada a una proteína PI3KCA mutante.
15. El método de una cualquiera de las reivindicaciones 10-14, en donde la secuencia de ADN diana comprende una mutación puntual T→C y/o an A→G que da como resultado:
- (i) una mutación de secuencia de aminoácidos en la proteína PI3KCA en comparación con la proteína PI3K no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la mutación puntual es una mutación A3140G que da como resultado una sustitución H1047R; o
 - (ii) una mutación de secuencia de aminoácidos en la proteína PSEN1 en comparación con la proteína PSEN1 no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína PSEN1 comprende una sustitución I143V provocada por una mutación puntual A→G en el codón 143 del gen PSEN1, opcionalmente en donde la mutación puntual en PSEN1 está asociada con enfermedad de Alzheimer, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 143 del gen PSEN1, corrigiendo así la mutación puntual A→G; o
 - (iii) una mutación de secuencia de aminoácidos en la proteína α-antitripsina en comparación con la proteína α-antitripsina no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína α-antitripsina comprende una sustitución L55P provocada por una mutación puntual T→C en el codón 55 del gen α-antitripsina, opcionalmente en donde la mutación puntual de α-antitripsina está asociada con enfermedad pulmonar obstructiva crónica (EPOC), opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 55 del gen α-antitripsina, corrigiendo así la mutación puntual T→C; o
 - (iv) una mutación de secuencia de aminoácidos en la proteína vWF en comparación con la proteína vWF no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína vWF comprende una sustitución C509A provocada por una mutación puntual T→C en el codón 509 del gen vWF, opcionalmente en donde la mutación puntual de vWF está asociada con enfermedad de von Willebrand, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 509 del gen vWF, corrigiendo así la mutación puntual T→C; o
 - (v) una mutación de secuencia de aminoácidos en la proteína caspasa-9 en comparación con la proteína caspasa-9 no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína caspasa-9 comprende una sustitución L197P provocada por una mutación puntual T→C en el codón 197 del gen caspasa-9, opcionalmente en donde la mutación puntual de caspasa-9 está asociada con neuroblastoma, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 197 del gen caspasa-9, corrigiendo así la mutación puntual T→C.
16. El método de una cualquiera de las reivindicaciones 10-15, en donde el método comprende además detectar la desaminación de la base nucleotídica, opcionalmente en donde la detección es por PCR.
17. La proteína de fusión de una cualquiera de las reivindicaciones 1-8 para su uso en un método de tratamiento de una enfermedad o trastorno, comprendiendo el método poner en contacto una molécula de ADN con
- (a) la proteína de fusión; y
 - (b) un ARN_g que dirige la proteína de fusión de (a) a una secuencia de ADN diana de la molécula de ADN; en donde la molécula de ADN se pone en contacto con la proteína de fusión y el ARN_g en una cantidad eficaz y en condiciones adecuadas para la desaminación de una base nucleotídica de la molécula de ADN,
- en donde la secuencia de ADN diana comprende una secuencia asociada a una enfermedad o trastorno, la desaminación de la base nucleotídica da como resultado una secuencia que no se asocia a una enfermedad o trastorno, y en donde el contacto es *in vivo* en un sujeto susceptible a tener, que tiene, o diagnosticado con una enfermedad o trastorno, opcionalmente en donde la secuencia de ADN diana comprende una mutación puntual asociada a una enfermedad o trastorno, y en donde la desaminación corrige la mutación puntual.
18. La proteína de fusión para su uso de la reivindicación 17, en donde la secuencia de ADN diana comprende una mutación puntual T→C o A→G asociada a una enfermedad o trastorno, y en donde la desaminación de la base C o G mutante da como resultado una secuencia que no se asocia a una enfermedad o trastorno.

19. La proteína de fusión para su uso de la reivindicación 17 o la reivindicación 18, en donde la secuencia asociada a la enfermedad o trastorno codifica una proteína, y en donde la desaminación introduce un codón de terminación en la secuencia asociada a la enfermedad o trastorno, dando como resultado una truncación de la proteína codificada.
- 5 20. La proteína de fusión para el uso de una cualquiera de las reivindicaciones 17-19, en donde la enfermedad o trastorno es fibrosis quística, fenilcetonuria, hiperqueratosis epidermolítica (EHK), enfermedad de Charcot-Marie-Toot de tipo 4J, neuroblastoma (NB), enfermedad de von Willebrand (vWD), miotonía congénita, amiloidosis renal hereditaria, cardiomiopatía dilatada (DCM), linfedema hereditario, enfermedad de Alzheimer familiar, enfermedad priónica, síndrome crónico, infantil, neurológico, cutáneo, articular (CINCA), miopatía relacionada con desmina (DRM), o una enfermedad neoplásica asociada a una proteína PI3KCA mutante.
- 10 21. La proteína de fusión para el uso de una cualquiera de las reivindicaciones 17-20, en donde la secuencia de ADN diana comprende una mutación puntual T→C y/o A→G que da como resultado:
- (i) una mutación de secuencia de aminoácidos en la proteína PI3KCA en comparación con la proteína PI3K no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la mutación puntual es una mutación A3140G que da como resultado una sustitución H1047R; o
- 15 (ii) una mutación de secuencia de aminoácidos en la proteína PSEN1 en comparación con la proteína PSEN1 no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína PSEN1 comprende una sustitución I143V provocada por una mutación puntual A→G en el codón 143 del gen PSEN1, opcionalmente en donde la mutación puntual en PSEN1 está asociada con enfermedad de Alzheimer, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 143 del gen PSEN1, corrigiendo así la mutación puntual A→G; o
- 20 (iii) una mutación de secuencia de aminoácidos en la proteína α-antitripsina en comparación con la proteína α-antitripsina no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína α-antitripsina comprende una sustitución L55P provocada por una mutación puntual T→C en el codón 55 del gen α-antitripsina, opcionalmente en donde la mutación puntual de α-antitripsina está asociada con enfermedad pulmonar obstructiva crónica (EPOC), opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 55 del gen α-antitripsina, corrigiendo así la mutación puntual T→C; o
- 25 (iv) una mutación de secuencia de aminoácidos en la proteína vWF en comparación con la proteína vWF no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína vWF comprende una sustitución C509A provocada por una mutación puntual T→C en el codón 509 del gen vWF, opcionalmente en donde la mutación puntual de vWF está asociada con enfermedad de von Willebrand, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 509 del gen vWF, corrigiendo así la mutación puntual T→C; o
- 30 (v) una mutación de secuencia de aminoácidos en la proteína caspasa-9 en comparación con la proteína caspasa-9 no mutante, y en donde el método da como resultado la desaminación de la base C o G mutante, opcionalmente en donde la proteína caspasa-9 comprende una sustitución L197P provocada por una mutación puntual T→C en el codón 197 del gen caspasa-9, opcionalmente en donde la mutación puntual de caspasa-9 está asociada con neuroblastoma, opcionalmente en donde el contacto da como resultado la desaminación del resto de citidina mutante en el codón 197 del gen caspasa-9, corrigiendo así la mutación puntual T→C.
- 35 40 22. La proteína de fusión para el uso de una cualquiera de las reivindicaciones 17-21, en donde el método comprende además detectar la desaminación de la base nucleotídica, opcionalmente en donde la detección es por PCR.

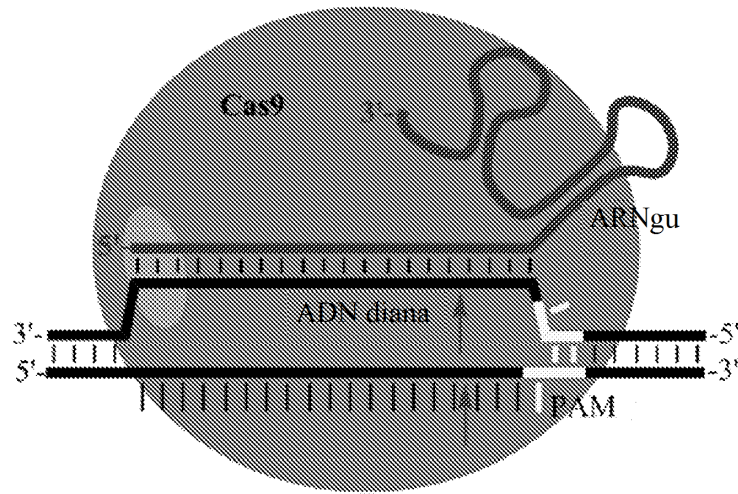


FIGURA 1



FIGURA 2

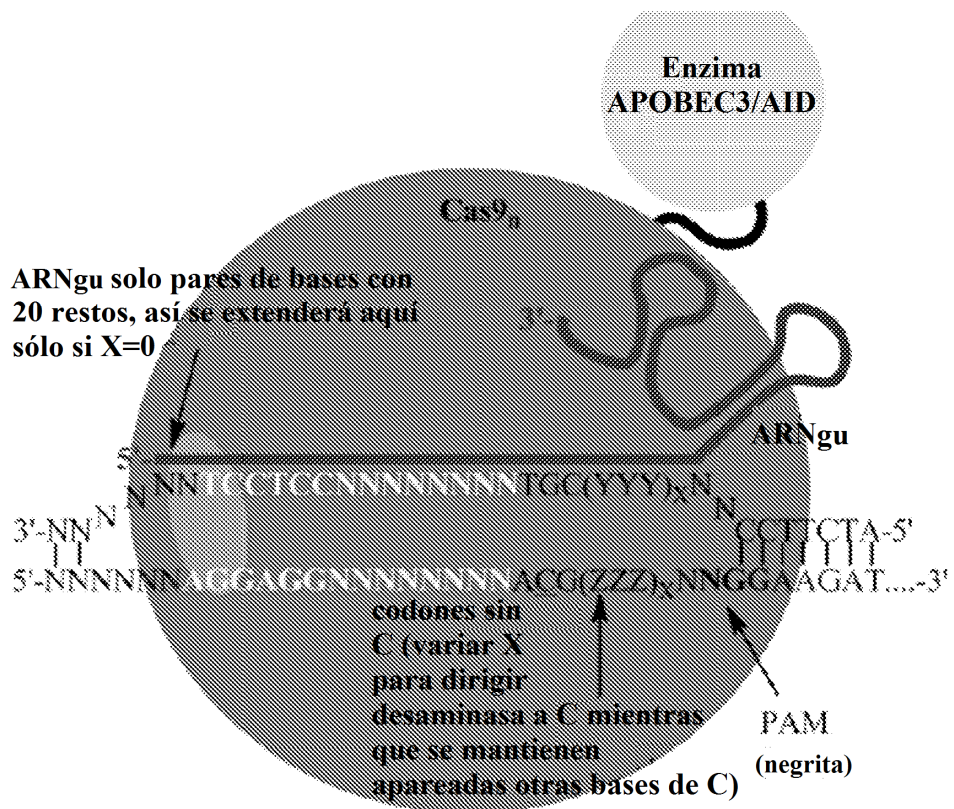


FIGURA 3

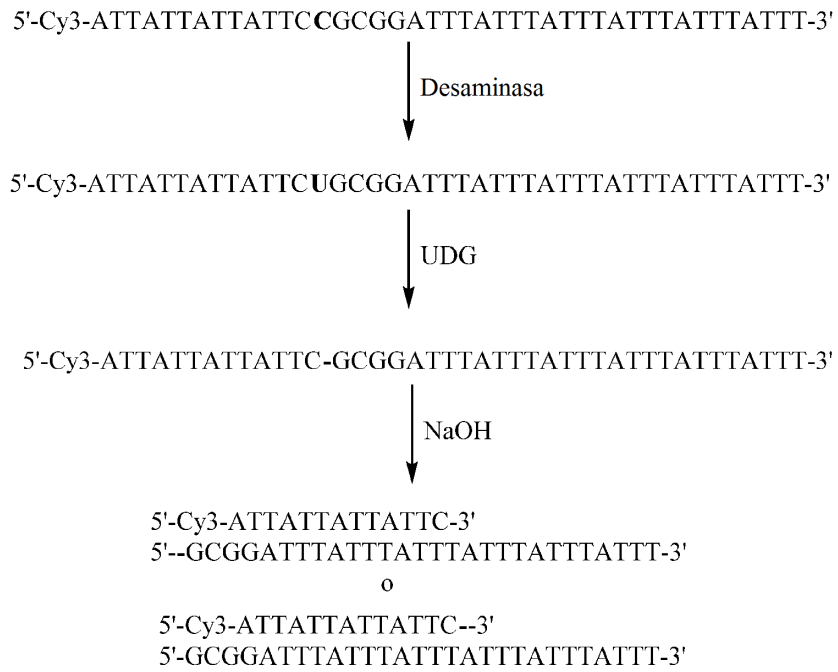


FIGURA 4

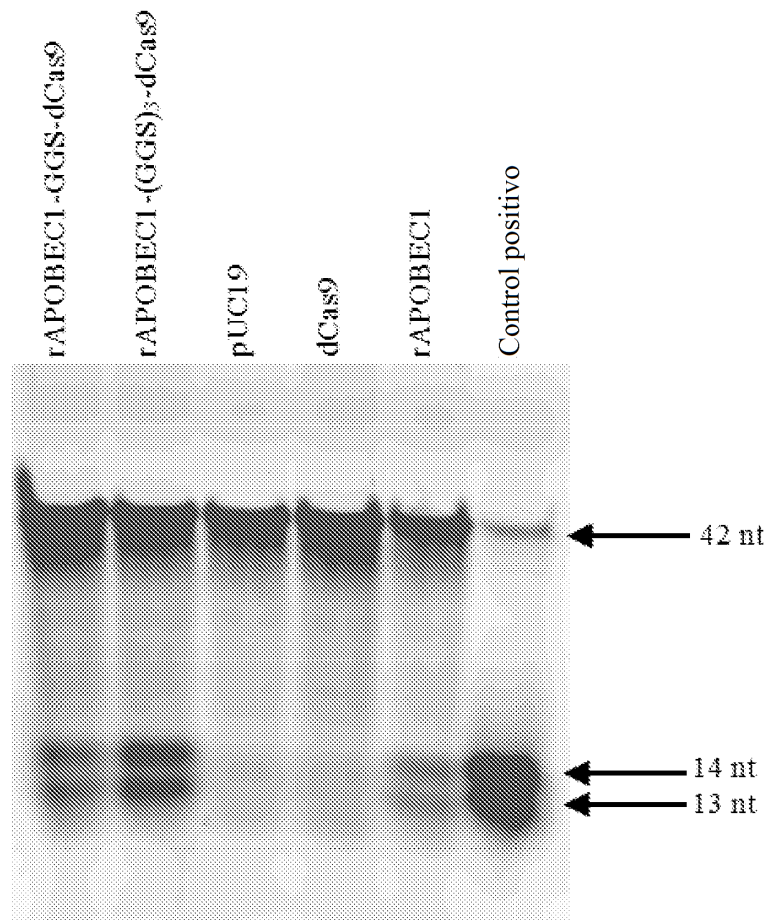


FIGURA 5