

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 992 912**

51 Int. Cl.:

C07K 14/195 (2006.01)
C12N 9/16 (2006.01)
C12N 9/22 (2006.01)
C12N 15/11 (2006.01)
C12N 15/62 (2006.01)
C12N 15/63 (2006.01)
C12N 15/79 (2006.01)
C12N 15/90 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **23.04.2018** **PCT/US2018/028919**
87 Fecha y número de publicación internacional: **25.10.2018** **WO18195545**
96 Fecha de presentación y número de la solicitud europea: **23.04.2018** **E 18788179 (2)**
97 Fecha y número de publicación de la concesión europea: **04.09.2024** **EP 3612551**

54 Título: **Variantes de CPF1 (CAS12A) con especificidad PAM alterada**

30 Prioridad:

21.04.2017 US 201762488426 P
11.01.2018 US 201862616066 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
19.12.2024

73 Titular/es:

THE GENERAL HOSPITAL CORPORATION
(100.0%)
55 Fruit Street
Boston, MA 02114, US

72 Inventor/es:

JOUNG, J. KEITH;
KLEINSTIVER, BENJAMIN y
SOUSA, ALEXANDER

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

ES 2 992 912 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 CPF1 (CAS12A) con especificidad PAM alterada

Reivindicación de prioridad

- 5 Esta solicitud reivindica el beneficio de la solicitud de patente estadounidense n.º de serie 62/488.426, presentada el 21 de abril de 2017, y 62/616.066, presentada el 11 de enero de 2018.

Investigación o desarrollo patrocinados por el gobierno federal

Esta invención se ha realizado con el apoyo del Gobierno en virtud de las subvenciones n.º GM105378 y GM118158 concedidas por los Institutos Nacionales de Salud. El Gobierno tiene ciertos derechos sobre la invención.

Campo técnico

- 10 La presente invención se refiere a secuencias de CRISPR modificadas mediante ingeniería de las proteínas *Prevotella* y *Francisella* 1 (Cpf1) de *Acidaminococcus* sp. BV3L6 (AsCpf1), que comprenden una secuencia que es al menos un 80% idéntica a la secuencia de aminoácidos de SEQ ID NO:2 con mutaciones en la posición E174R de la SEQ ID NO:2 con especificidades diana alteradas y mejoradas y sus usos.

Antecedentes

- 15 Las nucleasas CRISPR-Cas Cpf1 (también denominadas nucleasas Cas12a) han sido recientemente descritas como una alternativa a las nucleasas Cas9 para aplicaciones de edición del genoma (Zetsche et al. Cell 163, 759-771 (2015); Shmakov et al., Mol Cell. 2015 Nov 5;60(3):385-97; Kleinstiver et al., Nat Biotechnol. 2016 Ago;34(8):869-74; Kim et al., Nat Biotechnol. 2016 Ago;34(8):863-8). Las nucleasas Cpf1 poseen una serie de propiedades potencialmente ventajosas que incluyen, entre otras: reconocimiento de secuencias de motivos protoespaciadores-adyacentes (PAM) ricas en T, especificidades relativamente mayores en todo el genoma en células humanas que la Cas9 de *Streptococcus pyogenes* (SpCas9) de tipo salvaje, una actividad endorribonucleasa para procesar pre-ARNcr que simplifica la selección simultánea de múltiples sitios (multiplexación), actividad endonucleasa de ADN que genera un voladizo de ADN 5' (en lugar de una rotura roma de doble cadena como se observa con SpCas9), y escisión de la secuencia de ADN del protoespaciador en el extremo más distal del PAM (en comparación con la escisión en el extremo proximal del PAM del protoespaciador como se observa con SpCas9 y SaCas9). Hasta la fecha, los ortólogos de Cpf1 de *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* ND2006, y *Francisella tularensis* subsp. *Novicida* U112 (AsCpf1, LbCpf1, y FnCpf1 respectivamente), son los únicos ortólogos que han sido descritos que funcionan de forma robusta en células humanas. A pesar de estas capacidades, la adopción de las nucleasas Cpf1 para la edición del genoma ha sido menos rápida que la de SpCas9. Una posible razón podría ser el requisito de un PAM más largo que limita la selección a aproximadamente una vez cada 43 pbs de secuencia de ADN aleatoria, en comparación con una vez cada 8 pbs para SpCas9.

- 35 Aquí abordamos esta limitación del rango de selección utilizando un enfoque de ingeniería guiado por estructura para generar variantes de AsCpf1 no sólo con un rango de selección muy ampliado, sino también con actividades *on-target* sustancialmente mejoradas. Además de mejorar su potencia y versatilidad como nucleasas para la edición del genoma, demostramos que estas variantes pueden aprovecharse para otras aplicaciones, como la selección multiplex de nucleasas, la edición del epigenoma, la edición de bases C-a-T y la detección de ADN mediada por Cpf1, todo ello a niveles que antes no eran posibles con AsCpf1 de tipo salvaje.

Resumen

- 40 Aquí se describen, como parte de la divulgación, una serie de variantes de AsCpf1, FnCpf1 y LbCpf1 que reconocen una gama más amplia de PAMs que sus homólogos de tipo salvaje, aumentando así el rango de sitios que pueden ser objetivo de esta clase de nucleasas guiadas por ARN. Además, estas variantes se comportan mejor que las nucleasas Cpf1 de tipo salvaje a la hora de reconocer y modificar sitios diana que albergan PAMs TTTN canónicos. Las actividades mejoradas de las variantes aquí descritas mejoran las actividades de AsCpf1 para la edición del genoma, la edición del epigenoma, la edición de bases y la detección de ADN *in vitro*.

La invención se define en las reivindicaciones. La invención proporciona secuencias CRISPR aisladas de las proteínas *Prevotella* y *Francisella* 1 (Cpf1) de *Acidaminococcus* sp. BV3L6 (AsCpf1), que comprenden una secuencia que es al menos un 80% idéntica a la secuencia de aminoácidos SEQ ID NO:2 con mutaciones en las posiciones: E174R de la SEQ ID NO:2.

Las proteínas CpfI aisladas incluyen una mutación en E174R, opcionalmente con una o más mutaciones adicionales en S170R, K548, N551, T167, T539, S542, N552, M604, y/o K607. En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una mutación en S170R, opcionalmente con una o más mutaciones adicionales en E174R, K548, N551, T167, T539, S542, N552, M604, y/o K607.

- 5 En algunas realizaciones, las proteínas CpfI aisladas incluyen además una mutación en S542.

En algunas realizaciones, las proteínas CpfI aisladas incluyen además las mutaciones S542Q, S542K o S542R.

En algunas realizaciones, las proteínas CpfI aisladas incluyen una secuencia que es al menos un 95% idéntica a la secuencia de aminoácidos de SEQ ID NO:2.

- 10 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más de las siguientes mutaciones: E174R, S170R, K548R, S170K, E174K, T167A, T539K, T539R, K548V, N551R, N552R, M604A, K607Q, K607R, K607S, y/o K607H.

- 15 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen las siguientes mutaciones: S170R/E174R, E174R/K548R, S170R/K548R, E174R/S542R, S170R/S542R, E174R/S542R/K548R, E174R/N551R, S170R/N551R, S542R/K548R, S542R/N551R, S542R/N552R, K548R/N551R, S170R/S542R/K607R, E174R/S542R/K607R, E174R/S542R/K607H, E174R/S542R/K548R/N551R, S170R/S542R/K548V/N552R, E174R/S542R/K548V/N552R, S170R/S542R/K607R, o E174R/S542R/K607R de SEQ ID NO:2.

- 20 En algunas realizaciones, las proteínas CpfI aisladas incluyen una o más mutaciones que disminuyen la actividad de la nucleasa, por ej., seleccionadas del grupo que consta de mutaciones en D908, E993, R1226, D1235, y/o D1263, preferentemente D908A, E993A, R1226A, D1235A, y/o D1263A.

- 25 En algunas realizaciones, las proteínas CpfI aisladas incluyen además una mutación en uno o más de N282, N178, S186, N278, R301, T315, S376, N515, K523, K524, K603, K965, Q1013, Q1014, y/o K1054, preferentemente en N282, T315, N515, o N278, preferentemente donde la mutación aumenta la especificidad de la proteína. En algunas realizaciones, las proteínas CpfI aisladas incluyen además una mutación seleccionada del grupo que consiste en N282A, T315A, N515A, o N278A.

También se proporciona CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Lachnospiraceae bacterium ND2006* (LbCpfI), que no forma parte de la invención reivindicada, que comprende una secuencia que es al menos un 80% idéntica a la SEQ ID NO:11, con una o más de las siguientes posiciones: T152, D156, G532, y/o K538 de SEQ ID NO: 11.

- 30 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una secuencia que es al menos un 95% idéntica a la secuencia de aminoácidos de SEQ ID NO:11.

En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más de las siguientes mutaciones: T152R, T152K, D156R, D156K, G532R, y/o K538R.

- 35 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen las siguientes mutaciones: D156R/G532R/K538R.

En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más mutaciones que disminuyen la actividad de la nucleasa, por ej., seleccionadas del grupo que consta de mutaciones en D832, E925, R1138, D1148, y/o D1180, preferentemente D832A, E925A, R1138A, D1148A, y/o D1180A.

- 40 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen mutaciones en una o más de S202, N274, N278, K290, K367, K532, K609, K915, Q962, K963, K966, K1002 y/o S1003, preferentemente en las que la mutación aumenta la especificidad de la proteína. En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más de las siguientes mutaciones: S202A, N274A, N278A, K290A, K367A, K532A, K609A, K915A, Q962A, K963A, K966A, K1002A y/o S1003A.

45

También se proporciona CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Francisella tularensis* (FnCpfI), que no forma parte de la invención reivindicada, que comprende una secuencia que es al menos un 80% idéntica a la SEQ ID NO:4, con mutaciones en una o más de las siguientes posiciones: K180, E184, N607, K613, D616, N617, y/o K671 de la SEQ ID NO:4.

En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una secuencia que es al menos un 95% idéntica a la secuencia de aminoácidos de SEQ ID NO:4.

En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más de las siguientes mutaciones: K180R, E184R, N607R, K613R, K613V, D616R, N617R, K671H, y K671R.

- 5 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen las siguientes mutaciones: N607R/K613R, N607R/K613V, N607R/K613V/D616R, o N607R/K613R/D616R.

- 10 En algunas realizaciones que no forman parte de la invención reivindicada, las proteínas CpfI aisladas incluyen una o más mutaciones que disminuyen la actividad de la nucleasa, por ej., seleccionadas del grupo que consta de mutaciones en D917, E1006, R1218, D1227, y/o D1255, preferentemente D917A, E1006A, R1218A, D1227A, y/o D1255A.

También se proporcionan proteínas de fusión que comprenden las proteínas CpfI definidas en las reivindicaciones, fusionadas a un dominio funcional heterólogo, con un enlazador interviniente opcional, donde el enlazador no interfiere con la actividad de la proteína de fusión.

- 15 En algunas realizaciones, el dominio funcional heterólogo es un dominio de activación transcripcional, por ej., la fusión VP16 tetramérica denominada VP64, Rta, NF- κ B p65, o VPR (una proteína de fusión Rta, VP64, p65).

En algunas realizaciones, el dominio funcional heterólogo es un silenciador transcripcional o un dominio de represión transcripcional. En algunas realizaciones, el dominio de represión transcripcional es un dominio de caja asociado a Krueppel (KRAB), un dominio represor ERF (ERD) o un dominio de interacción mSin3A (SID). En algunas realizaciones, el silenciador transcripcional es la Proteína 1 de la Heterocromatina (HP1).

- 20 En algunas realizaciones, el dominio funcional heterólogo es una enzima que modifica el estado de metilación del ADN, por ej., un ADN metiltransferasa (DNMT) o una proteína TET, por ej., TET1.

En algunas realizaciones, el dominio funcional heterólogo es una enzima que modifica una subunidad de histona, por ej., una histona acetiltransferasa (HAT), histona desacetilasa (HDAC), histona metiltransferasa (HMT) o histona desmetilasa.

- 25 En algunas realizaciones, el dominio funcional heterólogo es una deaminasa que modifica bases de ADN de citosina, por ej., una citidina deaminasa de la familia de deaminasas de la enzima de edición de ARNm de la apolipoproteína B, polipéptido catalítico (APOBEC), incluyendo APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/E, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, citidina deaminasa inducida por activación (AID), citosina deaminasa 1 (CDA1) y CDA2, y citosina deaminasa que actúa sobre el ARNt (CDAT).

- 30 En algunas realizaciones, el dominio funcional heterólogo es una deaminasa que modifica bases de ADN de adenosina, por ej., la deaminasa es una adenosina deaminasa 1 (ADA1), ADA2; adenosina deaminasa que actúa sobre ARN 1 (ADAR1), ADAR2, ADAR3; adenosina deaminasa que actúa sobre ARNt 1 (ADAT1), ADAT2, ADAT3; y adenosina deaminasa específica de ARNt (TadA) natural o modificada por ingeniería.

- 35 En algunas realizaciones, el dominio funcional heterólogo es una enzima, dominio o péptido que inhibe o mejora la reparación endógena del ADN o las vías de reparación por escisión de bases (BER), por ej., el inhibidor de uracilo de ADN glicosilasa (UGI) que inhibe la escisión de uracilo ADN glicosilasa (UDG, también conocido como uracilo N-glicosilasa, o UNG) mediado por uracilo para iniciar BER; o proteínas de unión a extremos de ADN como Gam del bacteriófago Mu.

- 40 En algunas realizaciones, el dominio funcional heterólogo es una biomolécula, por ej., MS2, Csy4 o proteína N lambda.

En algunas realizaciones, el dominio funcional heterólogo es FokI.

También se proporcionan aquí ácidos nucleicos aislados que codifican las proteínas variantes de CpfI y proteínas de fusión como se define en las reivindicaciones.

- 45 Además, se proporcionan también vectores que comprenden los ácidos nucleicos según se define en las reivindicaciones. En algunas realizaciones, que no forman parte de la invención reivindicada, el ácido nucleico aislado codifica un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Acidaminococcus* sp. BV3L6 (AsCpfI), con mutaciones en una o más de las siguientes posiciones: T167, S170, E174, T539, K548, N551, N552, M604, y/o K607 de la SEQ ID NO:2 y está enlazado operablemente a uno o más dominios reguladores para expresar un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Acidaminococcus* sp. BV3L6

(AsCpfI), con mutaciones en una o más de las siguientes posiciones: T167, S170, E174, T539, K548, N551, N552, M604, y/o K607 de la SEQ ID NO:2.

En algunas realizaciones, que no forman parte de la invención reivindicada, el ácido nucleico aislado codifica un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de la *Lachnospiraceae bacterium ND2006* (LbCpfI), con mutaciones en una o más de las siguientes posiciones: T152, D156, G532, y/o K538 de SEQ ID NO: 11 y está operablemente enlazado a uno o más dominios reguladores para expresar un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de la *Lachnospiraceae bacterium ND2006* (LbCpfI), con mutaciones en una o más de las siguientes posiciones: T152, D156, G532, y/o K538 de SEQ ID NO:11.

En algunas realizaciones que no forman parte de la invención reivindicada, el ácido nucleico aislado codifica un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Francisella tularensis* (FnCpfI), que comprende una secuencia que es al menos un 80% idéntica a la SEQ ID NO:4, con mutaciones en una o más de las siguientes posiciones: K180, E184, N607, K613, D616, N617, y/o K671 de la SEQ ID NO:4 y está enlazado operablemente a uno o más dominios reguladores para expresar un CRISPR aislado de las proteínas *Prevotella* y *Francisella* 1 (CpfI) de *Francisella tularensis* (FnCpfI), que comprende una secuencia que es al menos un 80% idéntica a la SEQ ID NO:4, con mutaciones en una o más de las siguientes posiciones: K180, E184, N607, K613, D616, N617, y/o K671 de la SEQ ID NO:4.

También se proporcionan células huésped, preferentemente células huésped de mamífero, que comprenden los ácidos nucleicos según se define en las reivindicaciones, y opcionalmente expresan una proteína o proteína de fusión según se define en las reivindicaciones.

Además, se proporcionan aquí métodos *ex vivo* o *in vitro* para alterar el genoma de una célula, el método comprende expresar en la célula, o contactar la célula con, una proteína aislada o proteína de fusión según se define en las reivindicaciones, y al menos un ARN guía que tiene una región complementaria a una porción seleccionada del genoma de la célula, es decir, en la que la región complementaria es adyacente a una secuencia de PAM que corresponda a la proteína o proteína de fusión, por ej., como se muestra en la Tabla B. En algunas realizaciones, la proteína aislada o la proteína de fusión comprende una o más de una secuencia de localización nuclear, una secuencia peptídica de penetración celular y/o un marcador de afinidad. En algunas realizaciones, la célula es una célula madre. En algunas realizaciones, la célula es una célula madre embrionaria, una célula madre mesenquimal o una célula madre pluripotente inducida; se encuentra en un animal vivo o en un embrión. La invención también proporciona la proteína aislada o proteína de fusión según se define en las reivindicaciones para el uso en un método de alteración del genoma de una célula, el método comprende expresar en la célula, o contactar con la célula con, una proteína aislada o proteína de fusión de la invención, y al menos un ARN guía que tiene una región complementaria a una porción seleccionada del genoma de la célula.

También se proporcionan métodos *in vitro* o *ex vivo* de alteración de una molécula de ADN bicatenario (ADNs), el método comprende el contacto de la molécula de ADNs con una proteína aislada o proteína de fusión de la invención, y un ARN guía que tiene una región complementaria a una porción seleccionada de la molécula de ADNs, es decir, donde la región complementaria es adyacente a una secuencia de PAM que corresponde a la proteína o proteína de fusión, por ej., como se muestra en la Tabla B. La invención también proporciona la proteína aislada o proteína de fusión como se define en las reivindicaciones para el uso en un método de alteración de una molécula de ADN bicatenario (ADNs), el método comprende el contacto de la molécula de ADNs con una proteína aislada o proteína de fusión como se define en las reivindicaciones, y un ARN guía que tiene una región complementaria a una porción seleccionada de la molécula de ADNs, es decir, donde la región complementaria es adyacente a una secuencia de PAM que corresponde a la proteína o proteína de fusión.

También se proporcionan métodos para detectar un ADNss (monocatenario) o ADNs (bicatenario) diana *in vitro* en una muestra que no forma parte de la invención reivindicada. Los métodos incluyen poner en contacto la muestra con (i) la proteína aislada o proteína de fusión de las reivindicaciones 1-43, (ii) un ARN guía que tiene una región complementaria a una porción seleccionada de la molécula diana de ADNss o ADNs, y (iii) un ADN detector marcado, donde la proteína aislada o proteína de fusión escinde el ADN detector al unirse al ADNss o ADNs diana; y medir una señal detectable producida por la escisión del ADN detector marcado, detectando así el ADNss o ADNs diana.

En algunas realizaciones que no forman parte de la invención reivindicada, la medición del ADN detector marcado comprende la detección de una o más de una nanopartícula de oro, un fluoróforo, polarización de fluorescencia, transición/dispersión de fase coloidal, señales electroquímicas y señales basadas en semiconductores.

En algunas realizaciones que no forman parte de la invención reivindicada, el ADN detector marcado produce una cantidad de señal detectable antes de ser escindido, y la cantidad de señal detectable se reduce cuando el ADN detector marcado se escinde. En algunas realizaciones que no forman parte de la invención reivindicada, el ADN

detector marcado produce una primera señal detectable antes de ser escindido y una segunda señal detectable cuando el ADN detector marcado se ha escindido.

En algunas realizaciones que no forman parte de la invención reivindicada, el ADN detector marcado comprende un par quencher/fluor o un par FRET.

- 5 Salvo que se defina lo contrario, todos los términos técnicos y científicos aquí utilizados tienen el mismo significado que comúnmente entendería una persona con conocimientos ordinarios en la técnica a la que pertenece esta invención. También se describen aquí los métodos y materiales para el uso en la presente invención; asimismo pueden utilizarse otros métodos y materiales adecuados conocidos en la técnica. Los materiales, métodos y ejemplos son meramente ilustrativos y no pretenden ser limitativos. En caso de conflicto, prevalecerá la presente especificación, incluyendo las definiciones.

Otras características y ventajas de la invención se desprenderán de la siguiente descripción detallada y de las figuras, así como de las reivindicaciones.

Descripción de las ilustraciones

- 15 FIG. 1: Sitios diana representativos para las nucleasas Cpfl y SpCas9. AsCpfl y LbCpfl reconocen una secuencia PAM TTTN en el extremo 5' del espaciador del sitio diana, mientras que SpCas9 reconoce una secuencia de PAM NGG en el extremo 3' del espaciador. Secuencia de sentido: SEQ ID NO:1.

- FIG. 2: Perfiles de reconocimiento de PAM de las nucleasas AsCpfl y LbCpfl de tipo salvaje. Las capacidades de AsCpfl y LbCpfl de tipo salvaje para reconocer sitios diana en el ensayo de disrupción de EGFP en células humanas se determinaron utilizando ARNcr dirigido a sitios con PAM TTTN canónico o sitios con PAMs no canónicos con diferencias de una sola base. Se examinaron tres sitios para cada PAM no canónico, con la excepción de ATTN y TTAN debido a la falta de sitios diana en el gen reportero *EGFP*.

- FIG. 3: Perfiles de reconocimiento de PAM de AsCpfl de tipo salvaje para sitios con PAMs canónicos y PAMs no canónicos con diferencias de bases dobles y triples. La capacidad de AsCpfl para reconocer e interrumpir sitios diana en el ensayo de disrupción de EGFP en células humanas se determinó utilizando ARNcr dirigido a sitios con los PAM indicados. En la medida de lo posible, se examinaron tres sitios para cada PAM no canónico, con la excepción de ATTN y TTAN debido a la falta de sitios diana en el gen reportero *EGFP*. Barras de error, s.e.m. para $n = 3$, en caso contrario $n = 1$.

- FIG. 4: Perfiles de reconocimiento PAM de AsCpfl y LbCpfl para sitios PAM canónicos y no canónicos (diferencia de una sola base). Las capacidades de AsCpfl y LbCpfl de tipo salvaje para reconocer e interrumpir sitios diana endógenos en células humanas U2OS se determinaron mediante el ensayo T7E1 utilizando ARNcr dirigido a sitios con un PAM TTTN canónico o con un PAM no canónico con una sola diferencia de base. Donde fue posible, se examinaron tres sitios para cada PAM no canónico. Barras de error, m.e.s. para $n = 2$ o 3 , de lo contrario $n = 1$.

- FIG. 5A-5H: Alteración de la especificidad de reconocimiento de PAM AsCpfl por mutación racional de residuos físicamente proximales a bases de ADN PAM. Las actividades de AsCpfl de tipo salvaje y de las variantes que portaban mutaciones simples o dobles se evaluaron con ARNcr dirigido a sitios con PAM TTTN canónicos o no canónicos utilizando el ensayo de disrupción EGFP en células humanas o cuantificando las frecuencias de mutación indel inducidas en sitios de genes endógenos en células humanas U2OS. AsCpfl de tipo salvaje se comparó con: (A) variantes T167 / T539 utilizando ARNcr dirigido a sitios en EGFP; (B) variantes S170 y E174 utilizando ARNcr dirigido a sitios en EGFP; (C y D) variantes S542 utilizando ARNcr dirigido a sitios en EGFP (panel C) o sitios de genes humanos endógenos (panel D); (E) variantes N551 y N552 utilizando ARNcr dirigido a sitios en EGFP; (F) variantes K607 utilizando ARNcr dirigido a sitios en EGFP; (G y H) variantes S542 / K607 utilizando ARNcr dirigidos a sitios en EGFP (panel G) y sitios de genes humanos endógenos (panel H).

- FIG. 6A-6B: Perfiles de reconocimiento de PAM de AsCpfl de tipo salvaje y de la variante AsCpfl-S542R. Las actividades de estas nucleasas en el ensayo de disrupción de EGFP basado en células humanas fueron determinadas con ARNcr dirigido a sitios con un PAM TTTN canónico o un PAM no canónico con una sola diferencia de base (panel A) o diferencias de base dobles o triples (panel B). Cuando fue posible, se examinaron tres sitios para cada PAM no canónico, con la excepción de ATTN y TTAN debido a la falta de sitios diana suficientes en el gen reportero *EGFP* = 1.

- FIG. 7A-7B: Perfiles de reconocimiento de AsCpfl de tipo salvaje y de la variante AsCpfl-S542R en sitios diana de genes humanos endógenos. Se evaluó la capacidad de las nucleasas para mutagenizar sitios diana de genes endógenos en células U2OS humanas utilizando ARNcr dirigido a sitios con un PAM TTTN canónico o un PAM no

canónico con una sola diferencia de base (panel A) o diferencias de base dobles o triples (panel B). Donde fue posible, se examinaron tres sitios para cada PAM no canónico.

FIG. 8A-8C: Perfiles de reconocimiento de PAM de variantes de AsCpfl diseñadas racionalmente con mutaciones adicionales en residuos situados cerca de bases de ADN de PAM. En experimentos separados mostrados en los paneles A, B y C, se probaron sustituciones de aminoácidos individuales y combinaciones dobles o triples de sustituciones de aminoácidos utilizando el ensayo reporter EGFP basado en células humanas para evaluar sus capacidades para reconocer sitios diana con sitios PAM TTTN canónicos o no canónicos. FIG. 9A-9B: PAM de AsCpfl de tipo salvaje y variantes de AsCpfl. (A) La actividad de AsCpfl de tipo salvaje se comparó con las actividades de las variantes con una sola sustitución S542R o una combinación de las sustituciones E174R, S542R y K548R. Las actividades de estas nucleasas se probaron mediante un ensayo de disrupción de EGFP basado en células humanas utilizando ARNcr dirigido a sitios con un PAM TTTN canónico o PAMs con diferencias de una sola base. (B) La actividad de AsCpfl de tipo salvaje también se comparó con la variante E174R/S542R/K548R utilizando el ensayo reporter EGFP basado en células humanas con sitios diana de ARNcr con un PAM TTTN canónico o PAM con diferencias de base dobles o triples. Donde fue posible, se examinaron tres sitios para cada PAM no canónico. Barras de error, m.e.s. para $n = 2$ o 3 , de lo contrario $n = 1$.

FIG. 10A-10B: Comparación de los perfiles de reconocimiento de PAM de las variantes LbCpfl de tipo salvaje y AsCpfl. Se comparó la capacidad de LbCpfl de tipo salvaje para modificar sitios diana de genes humanos endógenos con la de la variante AsCpfl E174R/S542R/K548R. En este experimento se utilizó ARNcr dirigido a sitios con PAMs TTTN canónicos y PAMs con diferencias de una sola base (panel A), o PAMs con diferencias de dos o tres bases (panel B).

FIG. 11A-11C: Perfiles de reconocimiento de PAM de AsCpfl de tipo salvaje y diversas variantes de AsCpfl modificadas mediante ingeniería. La actividad de AsCpfl de tipo salvaje se comparó con las actividades de las variantes PAM de AsCpfl utilizando ARNcr dirigido a sitios con PAMs TTTN canónicos o PAMs con diferencias de una, dos o tres bases. Las comparaciones se realizaron (A) con el ensayo de disrupción de EGFP basado en células humanas o (B) evaluando las frecuencias de mutación (según el ensayo de endonucleasas T7EI) en los sitios diana de genes humanos endógenos. (C) Se compararon las actividades mutacionales de dos variantes de PAM AsCpfl utilizando ARNcr dirigido a sitios de genes humanos endógenos con PAMs TTTN canónicos o PAMs con diferencias de una, dos o tres bases.

FIG. 12A-12B: Perfiles de reconocimiento de PAM de LbCpfl de tipo salvaje y variantes de LbCpfl modificadas mediante ingeniería. (A) La actividad de LbCpfl de tipo salvaje y las variantes con diversas sustituciones de aminoácidos individuales se utilizó en el ensayo de disrupción de EGFP basado en células humanas con ARNcr dirigido a sitios con un PAM TTTN canónico o un PAM con una diferencia de una sola base. (B) La actividad de LbCpfl de tipo salvaje también se comparó con la variante LbCpfl1-D156R/G532R/K538R utilizando el ensayo de disrupción de EGFP basado en células humanas con ARNcr dirigido a sitios con un PAM TTTN canónico o un PAM con una diferencia de una sola base. $n = 1$.

FIG. 13A-13C: Perfiles de reconocimiento de PAM de FnCpfl de tipo salvaje y variantes de LbCpfl modificadas mediante ingeniería. (A) La actividad de FnCpfl de tipo salvaje utilizando el ensayo de disrupción de EGFP basado en células humanas con ARNcr dirigido a sitios que presentan un PAM TTN canónico o un PAM con una diferencia de una sola base. (B) La actividad de FnCpfl de tipo salvaje, AsCpfl y LbCpfl frente a sitios diana endógenos de células humanas con ARNcr dirigido a sitios con PAMs TTTN. (C) Comparación de la actividad del FnCpfl de tipo salvaje con las variantes PAM del FnCpfl modificadas mediante ingeniería utilizando el ensayo de disrupción de EGFP basado en células humanas con ARNcr dirigido a sitios que presentan un PAM TTTN canónico o un PAM con una sola diferencia de base. $n = 1$.

FIG. 14A-14C: Comparación de los perfiles de reconocimiento de PAM de las variantes de AsCpfl descritas en esta solicitud con diferentes variantes de AsCpfl divulgadas en otros trabajos. (A, B) Las actividades de varias de nuestras variantes de reconocimiento de PAM de AsCpfl modificadas mediante ingeniería se compararon con las variantes de reconocimiento de PAM S542R/K548V/N552R (panel A) y S542R/K607R (panel B) utilizando el ensayo de disrupción de EGFP basado en células humanas con ARNcr dirigido a sitios con PAMs TTTN canónicos o PAMs con diferencias de una o dos bases. (C) Se realizaron comparaciones adicionales de nuestras variantes AsCpfl con la variante S542R/K607R examinando las capacidades de estas nucleasas para mutagenizar sitios de genes humanos endógenos con ARNcr dirigido a sitios con PAMs TTTN canónicos o PAMs con diferencias de una, dos o tres bases.

FIG. 15A-15G: Ingeniería y caracterización de una variante de AsCas12a con rango de selección ampliado. (A), Modificación de sitios endógenos en células humanas por variantes de AsCas12a con sustituciones de aminoácidos. Actividades evaluadas mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n > 3$. (B), Perfiles de preferencia PAM para AsCas12a de tipo salvaje y la variante E174R/S542R/K548R, evaluados

mediante el ensayo de determinación de PAM (PAMDA). Las constantes de velocidad $\log_{10}$ (k) son la media de cuatro réplicas, dos frente a dos secuencias espaciadoras distintas (ver Fig. 21D). (C), Gráficos de actividad media para AsCas12a E174R/S542R y E174R/S542R/K548R en PAMs no canónicos, donde la línea negra representa la media de 12 a 20 sitios (puntos) para cada clase de PAM (ver también Figs. 23A, 23B y 23D). (D), Resumen de las actividades de AsCas12a de tipo salvaje, E174R/S542R y E174R/S542R/K548R en 20 sitios que codifican PAMs no canónicos, uno para cada PAM de las clases VTTN, TTCN y TATN (ver también las Fig. 19A, 23A y 23B; todos los sitios numerados '1'). (E), Gráficos de actividad media para AsCas12a, la variante E174R/S542R y eAsCas12a en PAMs TTTN, donde la línea negra representa la media de 5 a 8 sitios (puntos) para cada clase de PAM (ver Fig. 23G). (F), Superposición de los resúmenes de las actividades de células humanas y las constantes de velocidad (k) de PAMDA para varios PAMs seleccionables y no seleccionables con eAsCas12a (E174R/S542R/K548R). Se muestran gráficos de distribución de datos para las actividades de células humanas determinadas por el ensayo T7E1. Los PAMs de nivel 1 presentan un rango de selección medio superior al 20% en células humanas y un PAMDA k superior a 0,01, y los PAM que alcanzan un umbral modesto de rango de selección medio superior al 10% en células y un PAMDA k superior a 0,005 se consideran PAMs de nivel 2. (G), Cálculo de las mejoras en el rango de selección permitido por las variantes de AsCas12a en comparación con AsCas12a de tipo salvaje, determinado enumerando las secuencias PAM completas dentro de la característica de secuencia indicada y normalizando para el tamaño del elemento. TSS, sitio de inicio de la transcripción; las secuencias PAM seleccionables por cada variante de AsCas12a son: tipo salvaje, TTTV; eAsCas12a, ver el panel F y la Fig. 23I de datos ampliados para los niveles de PAM; RVR, TATV; RR, TYCV.

FIG. 16A-16E: Actividades mejoradas de las variantes de AsCas12a. (A), Cuantificación de las reacciones de escisión *in vitro* a lo largo del tiempo de los ortólogos y variantes de Cas2a en sustratos de plásmidos linealizados que codifican el sitio 1 de PAMDA diana, realizada a 37, 32 y 25 °C (paneles izquierdo, central y derecho, respectivamente). Las curvas se ajustaron utilizando una ecuación de decaimiento exponencial de una fase; las barras de error representan s.e.m. para $n = 3$. (B-D), Resúmenes de las actividades de las nucleasas AsCas12a de tipo salvaje y variantes en los sitios que codifican PAMs TTTN (panel B), PAMs TATN (panel C) y PAMs TYCN (panel D) (ver también las Figs. 24A-C, respectivamente). (E), Diagramas de dispersión de las constantes de velocidad determinadas por PAMDA para cada PAM NNNN para comparar las preferencias de PAM de las variantes AsCas12a (RVR a eRVR, panel izquierdo; RR a eRR, panel derecho). Las variantes codifican las siguientes sustituciones: eAsCas12a, E174R/S542R/K548R; RVR, S542R/K548V/N552R; eRVR, E174R/S542R/K548V/N552R; RR, S542R/K607R; eRR, E174R/S542R/K607R.

FIG. 17A-17H: Caracterización y mejora de la especificidad de eAsCas12a. (A), Perfiles de especificidad en todo el genoma realizados con la técnica GUIDE-seq para AsCas12a, eAsCas12a, y eAsCas12a-HF1 cada uno emparejado con sitios diana de ARNcr con PAMs TTTV. Se resaltan las posiciones discordantes en los sitios *off-target* (fuera del objetivo); los recuentos de lecturas de GUIDE-seq se muestran a la derecha de las secuencias; los círculos amarillos indican los sitios *off-target* que solo están respaldados por lecturas de GUIDE-seq asimétricas; los círculos verdes indican los sitios *off-target* identificados previamente para LbCas12a (Kleinstiver et al., Nat Biotechnol., 2016, 34:869-74); los nucleótidos alternativos en los PAMs no canónicos con PAMDA ks medio > 0,005 para eAsCas12a no están coloreados/resaltados como discordantes. SEQ ID NOs. 449-477, en orden de aparición. (B), Histograma del número de sitios *off-target* detectados por GUIDE-seq para las variantes de AsCas12a a partir de los sitios examinados en el panel A. (C), Diagrama de dispersión de las constantes de velocidad determinadas de PAMDA para cada PAM NNNN para comparar las preferencias de PAM de eAsCas12a y eAsCas12a-HF1. (D), Perfiles de especificidad en todo el genoma con la técnica GUIDE-seq para eAsCas12a y eAsCas12a-HF1 para sitios diana de ARNcr con PAMs no canónicos. Ilustraciones como las descritas para el panel a; eAsCas12a-HF1 no evaluado en CTTA-1, CTTC-2, o TATC-1. SEQ ID NOs. 478-530, en orden de aparición. (E), Histograma del número de sitios *off-target* detectados por GUIDE-seq para eAsCas12a y eAsCas12a-HF1 a partir de los sitios examinados en el panel a; na, no evaluado. (F), Ratio de eficiencia *off-target* calculada normalizando los recuentos de lecturas de GUIDE-seq *off-target* frente a los recuentos observados en el sitio *on-target* (en el objetivo). (G, H), Resumen de la actividad *on-target* de eAsCas12a y eAsCas12a-HF1 de tipo salvaje en los sitios que codifican PAMs TTTN (panel G) o PAMs no canónicos (panel H) (ver las Figs. 25I y 25J, respectivamente).

FIG. 18A-18K: Aplicaciones de eAsCas12a para la selección multiplex, la activación de genes y la edición de bases. (A-C), Comparación de las eficiencias de modificación *on-target* multiplex de AsCas12a, eAsCas12a y LbCas12a, cuando se programan con matrices de ARNcr dirigido a PAM TTTV que codifican 3 ARNcr separados expresados a partir de un promotor de la polimerasa III (U6, paneles A y B) o un promotor de la polimerasa II (CAG, panel C). Las actividades en tres loci distintos se evaluaron mediante el ensayo T7E1 utilizando las mismas muestras de ADN genómico; se muestran la media, s.e.m y los puntos de datos individuales para $n = 3$. (D), Evaluación de las eficiencias de edición cuando se utilizan plásmidos de ARNcr agrupados o matrices de ARNcr multiplex que expresan dos ARNcr dirigidos a loci genómicos cercanos (-100 pb). Actividades evaluadas mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n = 4$. (E-G), Activación de genes humanos endógenos con fusiones dCas12a- VPR (1.1) (ver Fig. 26A) utilizando grupos de tres ARNcr dirigidos a sitios PAM

canónicos (panel E) y sitios PAM no canónicos (paneles F y G). Las actividades evaluadas mediante RT-qPCR y la relación de cambio en el ARN se normalizaron a los niveles de *HPRT1*; se muestra la media, s.e.m y los puntos de datos individuales para triplicados técnicos de tres réplicas biológicas (n = 9). (H), Esquema de las construcciones del editor de bases (BE) dCas12a con distintas composiciones de NLS y enlazador. (I), Eficiencias de conversión de citosina a timina (C-a-T) dirigidas por dCas12a-BEs en ocho sitios diana diferentes, evaluadas mediante secuenciación profunda dirigida. Se examinó el porcentaje medio de edición de C-a-T de tres réplicas biológicas dentro de una ventana de -5 a +25; todas las Cs en esta ventana se resaltan en verde para cada sitio diana; la posición de la C dentro del sitio diana se indica debajo del mapa de calor. SEQ ID NOs. 531-538, en orden de aparición. (J), Resumen agregado de la eficiencia de edición C-a-T dentro de la secuencia espaciadora del sitio diana 20 nt con dCas12a-BEs en los ocho sitios diana. (K), Resumen de la relación de cambio en el porcentaje de lecturas de secuenciación que contienen mutaciones de inserción o delección (indels) para cada experimento dCas12a-BE (ocho sitios diana y tres réplicas), normalizado en relación con el porcentaje de indels observado en la muestra de control (LbBE1.4 y un plásmido U6 vacío). VPR, dominio sintético de activación VP64-p65- Rta (Chavez et al., Nat Methods., 2015, 12:326-8); NLS(sv), señal de localización nuclear de SV40; NLS(nuc), señal de localización nuclear de nucleoplasmina; rAPO1, APOBEC1 de rata; gs, enlazador peptídico de glicina-serina; UGI, inhibidor de uracilo glicosilasa.

FIG. 19A-19B: Actividades de ortólogos de Cas12a en células humanas. (A), Actividades de ortólogos de Cas12a dirigidos a sitios endógenos en células humanas con PAMs TTTN o VTTN. Modificación porcentual evaluada mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para n = 3. (B), Resumen de las actividades de los ortólogos de Cas12a frente a 24 sitios con secuencias PAM NTTN (se muestran las actividades medias de los datos del panel a).

FIG. 20A-20D: Ingeniería y caracterización de variantes de AsCas12a. (A), Representaciones esquemáticas y estructurales de Cas12a emparejado con un ARNcr e interactuando con un sitio diana putativo que codifica un PAM TTTV prototipo. En las representaciones estructurales, los residuos de aminoácidos proximales a las bases de ADN de PAM aparecen resaltados en verde; imágenes generadas a partir de PDBID:5B43 (Yamano et al., Cell. 2016 May 5; 165(4):949-62) visualizado en PyMOL (v 1.8.6.0). (B, C), Actividades de las variantes de AsCas12a con sustituciones de un solo aminoácido cuando se ensayan frente a sitios endógenos en células humanas con PAMs canónicos (panel B) o no canónicos (panel C). Modificación porcentual evaluada mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para n = 3. (D), Relación de cambio en las actividades medias de las variantes de AsCas12a en comparación con AsCas12a de tipo salvaje en sitios que presentan PAMs canónicos y no canónicos. Relación de cambio comparado con la actividad con AsCas12a de tipo salvaje calculado a partir de los datos de modificación porcentual de la Fig. 15A. FIG. 21A-21H: Optimización de un ensayo de caracterización *in vitro* de PAM. (A), Imágenes representativas en gel SDS-PAGE de ortólogos de Cas12a purificado y variantes de AsCas12a; s.m, marcador de tamaño en kDa. (B), Esquema del plásmido linealizado que presenta combinaciones de PAMs y espaciadores utilizados como sustratos para reacciones de escisión *in vitro*. SEQ ID NOs. 539-540. (C), Perfiles de reacción de escisión *in vitro* de AsCas12a de tipo salvaje (panel izquierdo) y variante E174R/S542R/K548R (panel derecho) sobre los sustratos ilustrados en el panel b. Las curvas se ajustaron utilizando una ecuación de decaimiento exponencial de una fase; las barras de error representan s.e.m para n = 3. (D), Esquema del ensayo de determinación de PAM (PAMDA). Las bibliotecas de plásmidos linealizados con 8 nucleótidos aleatorizados en lugar de PAM se sometieron a reacciones de escisión *in vitro* con complejos de ribonucleoproteína (RNP) de Cas12a. Los alícuotas de la reacción se detuvieron en varios puntos temporales, y posteriormente se utilizaron como plantilla para PCR. Se amplificaron y secuenciaron los sustratos que contenían PAMs incompletamente seleccionables para permitir la cuantificación de la tasa de agotamiento de PAM de la biblioteca de partida a lo largo del tiempo. (E), Correlación entre las constantes de velocidad PAMDA (k) a través de réplicas de AsCas12a de tipo salvaje (panel izquierdo) y la variante E174R/S542R/K548R (panel derecho). (F), Correlación entre las constantes de velocidad a partir de los valores medios de PAMDA en dos secuencias espaciadoras. (G), Histograma de las constantes de velocidad de PAMDA para AsCas12a de tipo salvaje y E174R/S542R/K548R. (H), Perfiles de agotamiento de los sustratos que codifican las secuencias PAM indicadas a lo largo del tiempo. Las curvas se ajustaron utilizando una ecuación de decaimiento exponencial de una fase; las barras de error representan s.e.m. para n = 4.

FIG. 22A-22C: Deconvolución de las especificidades de PAM de las variantes derivadas de eAsCas12a. (A), Perfiles de preferencia de PAM para AsCas12a de tipo salvaje, la variante E174R/S542R/K548R y todas las variantes intermedias de sustitución simple y doble, evaluadas mediante PAMDA. Las constantes de velocidad $\log_{10}$ (k) son la media de cuatro réplicas, dos frente a dos secuencias espaciadoras distintas (ver Fig. 21A-21H). (B), Comparación de los perfiles de preferencia de PAM de las variantes E174R/S542R y E174R/S542R/K548R en los 128 PAMs NNYN. (C), Alineación de los ortólogos de Cas12a con residuos importantes para alterar la preferencia de PAM en este estudio resaltados con un borde rojo. (SEQ ID NOs.:541-552).

FIG. 23A-23I: Evaluación del rango de selección mejorado de eAsCas12a. (A, B), Comparación de las actividades de E174R/S542R y E174R/S542R/K548R de AsCas12a en sitios endógenos en células humanas que presentan PAMs VTTN y TTCN no canónicos (panel a), o PAMs TATN (panel b). (C), Actividad de AsCas12a de tipo salvaje en sitios con PAMs TTCN o TATN. (D, E), Actividad de la variante E174R/S542R/K548R frente a sitios con PAMs TGTV (panel D) o sitios adicionales con varios PAMs no canónicos (panel E). (F), Correlación entre la constante de velocidad PAMDA y la modificación media en células humanas para PAMs probados en los paneles A-E. El recuadro sombreado en gris indica un umbral de la constante de velocidad de PAMDA arbitrario de 0,005 (o $10^{-2,25}$) que predice aproximadamente la actividad en células humanas. (G), Comparación de las actividades de AsCas12a de tipo salvaje, E174R/S542R y E174R/S542R/K548R en sitios con PAMs TTTN. (H), Resumen de las actividades medias de AsCas12a, la variante E174R/S542R y eAsCas12a en 26 sitios que codifican PAMs TTTN (ver también la Fig. 23G). (I), Resumen de PAMs seleccionables para eAsCas12a. Niveles de PAMs: 1, PAM de alta confianza (media $k > 0,01$, porcentaje medio modificado $> 13\%$); 2, PAM de confianza media (media $k > 0,005$, porcentaje medio modificado $> 10\%$); 3, PAM de baja actividad o discrepante (porcentaje medio modificado $< 10\%$ o discrepancia entre la media k y el porcentaje modificado). Para todos los paneles, porcentaje de modificación evaluado mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n > 3$.

FIG. 24A-24E: Actividades de las variantes de PAM Cas12a mejoradas. (A-C), Comparación de las actividades de modificación de sitios endógenos de las variantes de AsCas12a en sitios con PAMs TTTN (panel A), PAMs TATN (panel B) y PAMs TYCN (panel C). Modificación porcentual evaluada mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n = 3$. (D), Perfiles de preferencia de PAM para variantes AsCas12a RVR y RR originales y mejoradas evaluadas mediante PAMDA. Las constantes de velocidad $\log_{10}$ son la media de cuatro réplicas, dos frente a dos secuencias espaciadoras distintas (ver Fig. 21A- 21H). (E), Comparación de los perfiles de preferencia PAM de las variantes RVR/eRVR (panel superior) y RR/eRR (panel inferior) en los 128 PAMs NNYN. Las variantes AsCas12a codifican las siguientes sustituciones: eAsCas12a, E174R/S542R/K548R; RVR, S542R/K548V/N552R; eRVR, E174R/S542R/K548V/N552R; RR, S542R/K607R; eRR, E174R/S542R/K607R.

FIG. 25A-25K: Evaluación y mejora de las especificidades de AsCas12a y eAsCas12a. (A), Esquema del método GUIDE-seq. (B, C), Comparación de las actividades de mutagénesis *on-target* (panel b) e integración de marcadores dsODN mediante GUIDE-seq (panel c) de nucleasas AsCas12a para muestras de GUIDE-seq. Modificación porcentual e integración de marcadores evaluado mediante ensayos T7E1 y RFLP, respectivamente; media, s.e.m. y puntos de datos individuales mostrados para $n = 3$. (D), Ratio de integración de marcadores dsODN en GUIDE-seq en mutagénesis general de nucleasas AsCas12a; datos de los paneles b y c. (E), Actividades de AsCas12a de tipo salvaje o variantes con sustituciones únicas al usar ARNcr que coinciden perfectamente con el sitio *on-target*, o que codifican discordancias de un solo nucleótido. Porcentaje de modificación evaluado mediante ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n = 3$. SEQ ID NOs.553-554. (F), Actividades de variantes de eAsCas12a que presentan sustituciones de un solo aminoácido, evaluadas como en el panel e. Modificación porcentual evaluada mediante ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n = 3$. (G), Perfiles de preferencia de PAM de eAsCas12a y eAsCas12a-HF1 evaluados mediante PAMDA. Las constantes de velocidad de $\log_{10}$ son la media de cuatro réplicas, dos frente a dos secuencias espaciadoras distintas (ver Fig. 21A-21H). (H), Comparación de los perfiles de preferencia de PAM de eAsCas12a y eAsCas12a-HF1 en los 128 PAMs NNYN. (I, J), Evaluación de las actividades *on-target* de AsCas12a, eAsCas12a y eAsCas12a-HF1 en sitios diana que albergan PAMs TTTN (panel i) o PAMs VTTV, TATV y TTCV no canónicos (panel j). Modificación porcentual evaluada mediante el ensayo T7E1; media, s.e.m. y puntos de datos individuales mostrados para $n = 3$. (K), Curso temporal de las reacciones de escisión *in vitro* de los ortólogos de Cas12a y variantes en el sustrato PAMDA sitio 1, realizadas a 37, 32 y 25 °C (paneles izquierdo, central y derecho, respectivamente). Las curvas se ajustaron utilizando una ecuación de decaimiento exponencial de una fase; las barras de error representan s.e.m. para $n = 3$. Las variantes de AsCas12a codifican las siguientes sustituciones: eAsCas12a, E174R/S542R/K548R; eAsCas12a-HF1, E174R/N282A/S542R/K548R.

FIG. 26A-26F: Activación génica con fusiones de Cas12a. (A), Esquema de las fusiones de dominios de activación de VPR a ortólogos y variantes de Cas12a (dCas12a) Dnasa inactivo. (B), Ilustración de la ventana de secuencia que abarca aproximadamente 700 pb en sentido ascendente del sitio de partida de la transcripción VEGFA (TSS), con los sitios diana para SpCas9 y Cas12a indicados. (C, D), Comparación de las actividades de las arquitecturas dCas12a-VPR y dSpCas9-VPR (utilizando pares de ARNcr o ARNsg, respectivamente); el ARNcr se dirigió a sitios con PAMs TTTV (panel C) o PAMs TTCV (panel D) en el promotor de VEGFA. Actividades evaluadas a través de los cambios en la producción de VEGFA en comparación con una transfección de control conteniendo dAs-VPR(1.3) y un plásmido de ARNcr simulado; media, s.e.m. y puntos de datos individuales mostrados para $n = 4$. (E, F), Activación de VEGFA por las proteínas de fusión dCas12a-VPR(1.1) o dSpCas9-VPR utilizando grupos de tres o dos (paneles e y f, respectivamente) ARNcr o ARNsg a través de un rango de sitios con PAMs canónicos y no canónicos para las fusiones dCas12a-VPR; media, s.e.m., y puntos de datos individuales mostrados para $n > 3$. VPR, dominio sintético de activación VP64-p65-Rta (Chavez et al., Nat Methods., 2015, 12:326-8); NLS(sv), señal

de localización nuclear de SV40; NLS(nuc), señal de localización nuclear de nucleoplasmina; HA, marcador de hemaglutinina de gripe humana; gs, enlazador peptídico de glicina-serina.

FIG. 27A-D: Edición de bases con Cas12a. (A), Relación de cambio en la edición de C-a-T comparado con el control no tratado en todas las Cs en los espaciadores de 20 nt de 8 sitios diana. (B), Influencia de la identidad de la base precedente (5') en la conversión de citosina en timina (C-a-T). La eficiencia de la edición de C-a-T en ocho sitios diana (ver Fig. 18I) se traza para todas las Cs en la ventana que abarca la región de -14 a +30 de cada sitio diana (10 nt adicionales en sentido ascendente de PAM 4 nt y 10 nt en sentido descendente de la secuencia espaciadora de 20 nt). (C), Análisis de la pureza de edición en seis citosinas seleccionadas en cinco sitios diana. La fracción de cada identidad no-C se traza sobre la suma de todas las ocurrencias no-C en esa posición para cada construcción BE. (D), Se calcularon las actividades de mutación por inserción o delección (indel) de Cas12a-BEs para cada par BE/ARNcr mediante la determinación del porcentaje de alelos que codifican un indel dentro de la ventana de -14 a +30, sin contar los alelos con sustituciones únicamente.

FIG. 28A-28B: Detección de ADN con AsCas12 y eAsCas12a. (A), Detección temporal de ADN con AsCas12a de tipo salvaje mediante fluorescencia de molécula reporter inducida por DNasa. Actividades evaluadas cuando se programan con diferentes sustratos activos e inactivos. (B), Actividades de detección de ADN de eAsCas12a y eAsCas12a-HF 1 (paneles superior e inferior, respectivamente) a lo largo del tiempo. Las actividades se evaluaron cuando se programaron con sustratos con PAM canónicos, PAM no canónicos y secuencias no seleccionables. Las mediciones de fluorescencia se realizaron cada 60 segundos durante tres horas con $\lambda_{ex} = 485 \text{ nm}$ y $\lambda_{em} = 528 \text{ nm}$.

Descripción detallada

Las enzimas Cpf1 caracterizadas hasta la fecha reconocen los PAMs ricos en T que están situados a 5' de la secuencia espaciadora (Fig. 1). Tanto AsCpf1 como LbCpf1 han reconocido PAM de la forma TTTN, pero prefieren firmemente TTTV (donde V = A, C o G). Se espera que una secuencia PAM TTTV ocurra aproximadamente una vez cada 43 bases de ADN aleatorio, lo que limita potencialmente el rango de selección (y la utilidad) de AsCpf1 y LbCpf1 para la edición del genoma. El rango de selección de las nucleasas modificadas mediante ingeniería es especialmente importante para aplicaciones que requieren una selección o colocación precisa de la rotura de la doble cadena (DSB) del ADN, entre otros: 1) generación de mutaciones de inserción o delección de indels en pequeños elementos genéticos como: marcos abiertos de lectura (ORF) cortos, sitios de unión de factor de transcripción, micropéptidos, ARNm, etc.; 2) la reparación dirigida por homología (HDR), en la que la proximidad del DSB al cambio de secuencia deseado puede influir enormemente en la eficacia de la reparación, 3) edición alelo-específica conseguida colocando la variación SNP dentro del protoespaciador o PAM; 4) generar delecciones genómicas de longitud definida o translocaciones mediante la introducción de pares de DSB; 5) realizar mutagénesis de saturación de genes o elementos reguladores de genes; y 6) usar nucleasas guiadas por ARN modificado mediante ingeniería o nicasas fusionadas a enzimas modificadoras del ADN para realizar la edición de bases. Dadas estas claras ventajas de un mayor rango de selección, intentamos alterar o relajar las especificidades PAM de AsCpf1 y LbCpf1 para mejorar sus capacidades de reconocer un rango más diverso de secuencias de ADN.

Aquí demostramos que las sustituciones en o cerca de residuos de aminoácidos proximales de PAM pueden alterar las preferencias de PAM tanto de AsCpf1 como de LbCpf1, generando variantes de estas nucleasas que pueden reconocer secuencias de PAM no cognadas, aumentando así el rango de selección de esta plataforma. Estas variantes CRISPR-Cas12a modificadas mediante ingeniería han mejorado drásticamente sus propiedades, mostrando simultáneamente un rango de selección y una actividad de selección mejorada. Hasta donde sabemos, esta es la primera descripción de sustituciones de aminoácidos que pueden mejorar la actividad *on-target* de una nucleasa CRISPR. Las propiedades mejoradas de eAsCas12a ofrecen importantes ventajas sobre los ortólogos y variantes de Cas12a disponibles en la actualidad, ya que presentan una mejora de más de 8 veces el rango de selección, al tiempo que permiten aplicaciones más potentes de edición multiplex, activación génica, detección de ADN y edición de bases con eficiencias que antes eran inalcanzables con AsCas12a de tipo salvaje. El desarrollo de las tecnologías de edición de bases eAsCas12a amplía el alcance de las bases seleccionables en el genoma y lo hace con escasa evidencia de mutaciones indel colaterales. Es importante destacar que el rango de selección de eAsCas12a es comparable al de las nucleasas SpCas9 modificadas mediante ingeniería descritas anteriormente, lo que proporciona una mayor densidad de sitios diana para aplicaciones Cas12a que requieren un reconocimiento más amplio de PAM (por ej., para la selección dentro de ventanas genómicas definidas o pequeñas, edición multiplex del genoma o epigenoma, screens de ARNcr segmentado codificante o no codificante, o cuando se conciben screens de bibliotecas combinatorias complejas). Las propiedades mejoradas de las variantes aquí descritas, incluyendo eAsCas12a, las convierten en algunas de las enzimas Cas12a más activas y más ampliamente focalizables descritas hasta la fecha.

Cpf1

Los sistemas de repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) codifican endonucleasas guiadas por ARN que son esenciales para la inmunidad adaptativa bacteriana (Wright et al., Cell 164, 29-44 (2016)). Las nucleasas asociadas a CRISPR (Cas) pueden programarse fácilmente para escindir secuencias de ADN diana para la edición del genoma en diversos organismos²⁻⁵. Una clase de estas nucleasas, denominadas proteínas Cas9, forman un complejo con dos ARN cortos: un ARNcr y un ARNcr transactivador (ARNtracr)^{7,8}. El ortólogo de Cas9 más utilizado, SpCas9, utiliza un ARNcr que tiene 20 nucleótidos (nt) en su extremo 5' que son complementarios a la región "protoespaciadora" del sitio de ADN diana. La escisión eficiente también requiere que SpCas9 reconozca un motivo adyacente del protoespaciador (PAM). El ARNcr y el ARNtracr suelen combinarse en un único ARN guía (ARNg)^{7, 9-11} de ~100 nt que dirige la actividad de escisión del ADN de SpCas9. Las especificidades genómicas de las nucleasas SpCas9 emparejadas con diferentes ARNg se han caracterizado utilizando muchos enfoques diferentes¹²⁻¹⁵. También se han modificado mediante ingeniería variantes de SpCas9 con especificidades sustancialmente mejoradas en todo el genoma^{16, 17}.

Recientemente, se ha identificado una proteína Cas denominada Cpf1 que también puede programarse para escindir secuencias de ADN diana^{1, 18-20}. A diferencia de SpCas9, Cpf1 requiere un único ARNcr de 42 nt, que tiene 23 nt en su extremo 3' que son complementarios al protoespaciador de la secuencia de ADN diana¹. Además, mientras que SpCas9 reconoce una secuencia PAM NGG que se encuentra a 3' del protoespaciador, AsCpf1 y LbCpf1 reconocen PAMs TTTN que se sitúan a 5' del protoespaciador¹. Los primeros experimentos con AsCpf1 y LbCpf1 demostraron que estas nucleasas pueden programarse para editar sitios diana en células humanas¹ pero sólo se probaron en un pequeño número de sitios. Estudios recientes han demostrado que tanto AsCpf1 como LbCpf1 poseen una sólida actividad *on-target* y una elevada especificidad genómica en células humanas (ver, por ej., Kleinstiver & Tsai et al., Nature Biotechnology 2016; y Kim et al., Nat Biotechnol. 2016). Ver también US20160208243.

Los presentes hallazgos respaldan las variantes modificadas por ingeniería AsCpf1 y LbCpf1, denominadas colectivamente en este documento "variantes" o "las variantes".

Todas las variantes aquí descritas pueden incorporarse rápidamente a vectores existentes y ampliamente utilizados, por ej., mediante mutagénesis simple dirigida al sitio.

Así pues, en esta especificación se proporcionan variantes de AsCpf1. La secuencia de la proteína de tipo salvaje AsCpf1 es la siguiente:

AsCpf1 - Proteína asociada a CRISPR tipo V Cpf1 [Acidaminococcus sp. BV3L6], Secuencia de referencia NCBI:

WP_021736722.1

```

1 MTQFEGFTNL YQVSKTLRFE LIPQCKTLRH IQEQGFIEED KARNDRYREL
KPIIDRIYKT
51 YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHXYF
IGRTINKLTDN
121 INKRHAELIYR CLFKAELFNG KVLKQLGTVI TTEHENALLR SPDRKTTYFS
GFYENRKNVF
181 SAEDISTALF HRIVQDNFFK FRENCHIFTR LITAVPSLRE HEENVKRAIG
IFVTSIEEV
241 PSFFFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDK
TAHILASLPH
301 REIPLEKQIL SDRNTLSFIL BEFKSDDEVI QSFCKYKTLI RNENVLETAE
ALFNELNSID
361 LTHIFTSRKK LETISSALCD HWDITRNALY EBRISLITGK ITKSAKERVQ
RSLKHEDINL
421 QEIIISAAGKE LSEAFKQETS EILSHAHAAI DQPLPTTLKK QEENEILKSK
LDSLLGLYHL
481 LDWEFVDESN EVOFEFSAPL TGIKLEMEPS LSFYKAPNY ATKKEYSVEK
FELNFQMPETL
541 ASGWGVNNEK NNGAILFVKN GLYVLGIMPK QKGRYKALSF EPFEKTSEGF
DKMYDYFED
601 AARMIFKCSY QLKAVTAPFQ THTTFILLEN NFIEPLEITK EYGLNNPEK
EPKKEQTAYA
661 KKTGQKQYR EALCKWIDET RFLSKYTTT TSIDLSSLRP SSQYKDLGEX
YAEINELLNH
721 ISFQPIAERE IMDAVETGKL YLFQIYNRDF AKGHHGKPNL HTLYWTGLFS
PENLAKTSIK
781 LNSQAELEYR PKSRMKPMH RLGERMLNRK LKQCKTPIFD TLYQELYDYV
NHRLSHOLSO
841 EAKALLPNVI TKEVSHELK DRPFSDEFF PHVEITLNYQ AANSFSKFNQ
PVNAYLKEMP
901 EPTTIGIDRG ERNLIYITVI DSTGRILEQR SENTIQQFDY QKKLDNREKE
RVAARQAWSV
961 VGTIRDLKQG YLSQVINEIV DLMHYQAVV VLENLNFGEK SKRTGIAEKA
VYQOFKMLI
1021 DKLNCILVLEK YPAEKVGSVL NPYQLTDQFT SFARMGTQSG ELEYVFAFYT
SKIDFLTGFV
1081 DPFVWETIKN HESRKHFELEG FFLHYDVNT GDFILHFKMN RNLSFQSGLF
GEMFAWDIVF
1141 EKNETQFTAK GTPFIAGKRI VPVIENHRET GRYRDLYPAN ELIALLEEK
IVFRDGSNIL
1201 EKLENDSSH AIDTMVALIR SVLQMRNSNA ATGEDYINSE VPDLNGVCFD
SRFQNFEPW
1261 DADANGAYHI ALKGQELINH LKESKDLKLQ NGISNQDWLA YIQELRN (SEQ ID
NO:2)

```

- La variante AsCpf1 de la invención se define en las reivindicaciones. Las variantes de AsCpf1 aquí descritas pueden incluir la secuencia de aminoácidos de SEQ ID NO:2, por ej., al menos comprendiendo los aminoácidos 1-1307 de la SEQ ID NO:2, con mutaciones (es decir, sustitución del aminoácido nativo por un aminoácido diferente, por ej., alanina, glicina o serina (excepto cuando el aminoácido nativo es serina)), en una o más posiciones de la Tabla 1, por ej., en las siguientes posiciones: T167, S170, E174, T539, K548, N551, N552, M604, y/o K607 de SEQ ID NO:2 (o en posiciones análogas a las mismas, por ej., de SEQ ID NO:9). En algunas realizaciones de la divulgación, las variantes de AsCpf1 son al menos un 80%, por ej., al menos un 85%, 90% o 95% idénticas a la secuencia de aminoácidos de SEQ ID NO:2, por ej., tienen diferencias de hasta un 5%, 10%, 15% o 20% de los residuos de SEQ ID NO:2 sustituidos, por ej., con mutaciones conservadoras, además de las mutaciones aquí descritas. En realizaciones preferidas, la variante retiene la actividad deseada del progenitor, por ej., la actividad nucleasa (excepto cuando el progenitor es una nicasa o una Cpf1 muerta), y/o la capacidad de interactuar con un ARN guía y ADN diana).
- También se proporcionan variantes de LbCpf1 que no forman parte de la invención reivindicada. La secuencia de la proteína de tipo salvaje LbCpf1 es la siguiente:

LbCpf1 - Proteína asociada a CRISPR de tipo V Cpf1 [bacteria Lachnospiraceae ND2006], GenBank N.º de acceso WP_051666128.1

```

1 MLKNVGIIDPL DVERGRNMS KLERETNCTG LSXTLRFPAI PVGKTQENID
NERLLVEDER
61 PAEDYKGVKR LLDRYYLSFI NDVLHSIRLK NLNYYISLFR RKTREKENK
EENLEINLR
121 RELAKAFKGN EGYKSLFKD IETILPEFL DDKDEIALVN SFNGFTTAPT
GFFDMAENMF
181 GEEAKTSIA FRCINENLTP YISNMDIFEK VOAIFPKHEV QEIREKILNS
DYDVEOFFEG
241 EEFNFVLTQE GILVYNALIG GFVTESGEKI KGLNEYIQLY NQHTKQKLEK
FKFLYKQVLS
301 DRESLSFYGE GYTSDEEVLE VTRNLAKNS EIFSSIKKLE KLFKNFDEYS
SAGIFVKNGF
361 AISTISKDIF GENNVIAKDW NAEDDIDMLK KNAVVTENYE ODRKNSPKFI
GFSLEQLQE
421 YADADLSVVE KLEIITIQKV DEIKVYGES ENLEDAEVL EKSLEKNDV
VAIMEIDLES
481 VKSPENYIKA FFGCKETMR DESFYGDFVL AYDILLKVDH IYDAIRNYVT
QPFYSKDMFK
541 LYFQNPQFMG GWKDKETDY KATILRYGSK YLAIMDKRY AKCLQRIDED
DVNGNYEKIN
601 YKLLFGENKM LFKVFFSKKW MAYYNFSEDI QRIYKNGTEK NGDMFHLNDC
HKLIDFEKDS
661 ISRYPKWENA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASHKEVDK
IVVEGRLYMF
721 QIYNKDFSDK SHGTFNLHTM YFKLLFDENN HQIPLSGGA ELEMPRASLK
KEELVVMFAN
781 SPIANNEDN PKNTTTLSDY VYKDKFSED QYELHIPIAI NKOPFNIPKI
NTEVVEILNN
841 DENNYVIGID RGERNLLYIV VVDSKGNIVE QYSLNETINN ENGIRIKTDY
HSLLOKKEKE
901 REEARQNWTS IENIKELKAG YISQVVKIC ELVEKYDAVI ALEDLNSGFK
NSRVIVVEKQV
961 TQKFERMLID KLNIMVDNKS NPCATGGALK GKQITHKFS EKMSTQNGF
IFVIRAWLTS
1021 KIDESTGPNV LLKTKYTSIA DSKEFISSEF RIMYVEEDI EFALDYKMF
SRGDADYLRK
1081 WRLYSYGNRI RIFRNPKKNV VFDWEVCLT SAYKELFNKY GINYQQGDIR
ALLCEQSDRA
1141 FYSSPMALMS LMDQMRNSIT SRDIPVDFIS PVKNSGIFTY GSKNYEAQEN
ALLFKNAVAN

```

1201 GAYNIARKVL WAISQPMKAE DEKLDKXIA ISNKEWLEYA QTSVEN (SEQ ID NO:3)

Las variantes de LbCpfI aquí descritas pueden incluir la secuencia de aminoácidos de la SEQ ID NO:3, por ej., al menos comprendiendo aminoácidos 23-1246 de la SEQ ID NO:3, con mutaciones (es decir, sustitución del aminoácido nativo por un aminoácido diferente, por ej., alanina, glicina o serina), en una o más de las siguientes posiciones: T152, D156, G532, y/o K538 de SEQ ID NO: 11 (o en posiciones análogas, por ej., T170, D174, G550 y/o K556 de SEQ ID NO:3); los aminoácidos 19-1246 de SEQ ID NO:3 son idénticos a los aminoácidos 1-1228 de SEQ ID NO: 11 (aminoácidos 1-1228 de SEQ ID NO:11 también se denominan aquí LbCPF1 (-18)). En algunas realizaciones que no forman parte de la invención reivindicada, las variantes de LbCpfI son al menos un 80%, por ej., al menos un 85%, 90% o 95% idénticas a la secuencia de aminoácidos de SEQ ID NO:3, por ej., tienen diferencias de hasta un 5%, 10%, 15% o 20% de los residuos de SEQ ID NO:3 sustituidos, por ej., con mutaciones conservadoras, además de las mutaciones aquí descritas. En realizaciones preferidas que no forman parte de la invención reivindicada, la variante retiene la actividad deseada del progenitor, por ej., la actividad nucleasa (excepto cuando el progenitor es una nicasa o una CpfI muerta), y/o la capacidad de interactuar con un ARN guía y ADN diana). La versión de LbCpfI utilizada en los presentes ejemplos de trabajo comienza en el motivo MSKLEK, omitiendo los primeros 18 aminoácidos encuadrados anteriormente como se describe en Zetsche et al. Cell 163, 759-771 (2015).

También se proporcionan variantes de FnCpfI que no forman parte de la invención reivindicada. La secuencia de la proteína de tipo salvaje FnCpfI es la siguiente:

FnCpf1— Proteína asociada a CRISPR de tipo V Cpf1 [Francisella tularensis], GenBank N.º de acceso WP_003040289.1

1 MSIXQEIVNN YLSNTLRFE LIPQGNLTEN IKARGLILDD EKPADYKKA
KQIIDKHQF
61 FTEELSSVC ISEDLQNVY DVYFKLRKSD DDLQKDFKS AKDTIKKQIS
EXKDEKFK
121 NLPNQNLIDA HKQESDLIL WLKQSKDNGI ELFKANSBIT DIDEALBIK
SFRGWTTYPK
181 GFENKKNVY SSKDIPTSH TRIVEDNLPE FLENKAKYES LHKAFKAIN
YEQIKKDLAE
241 ELTFDIOYKT SEVNGRVFSL DEVFEIANFN NYLNQSGITK FNTIGGEFV
NGENTKRGSI
301 NEYINLYSQQ INDKTLKRYK MSVLEKQILS DTEKSFVID KLEDDSTVYT
TMQSPYEQIA
361 AKTVEEKSI KETLSLLEDD LKAQKLDLSK IYPMNDKSLT DLSQQVEDDY
SVIGTAVLEK
421 ITQQIAPKNL DNFSEKQEL IANKTEKAKY LSLETIKIAL BEFNKHROID
KQCFEEILA

481 NFAAEMIFD EIAQKKNLIA QISIKYQNGG KKDLLQASAE DPKAKRDLI
DQTNLHLHL
541 KIFHISQSED KANILDKDH FYLVFECCYF ELANIVELYN KIRNYITQKE
YSEKFKLNF
601 ENSTLANGWD KKKEPDNTAI LPIKDDKYVL GVMKKKNNKI EDORAKERNK
GEGYKKEVYK
661 LLEFANKMLP KVFPSAKSIK FYNPSEDILA IRNHSTHTEN GSPQKGYEKF
EFNIEDCKRF
721 IDFYKQSISK HPEWKDFGFR TSSTORYNSI DEFYREVENQ GYKLTFFNIS
ESYIDSVVNQ
781 GKLYLFQIYN KDFSAVSKGR PNLHTLYWKA LPDEPNLQDV VYKLNCEARI
FYKQSIPEK
841 ITHAKKAAIA NKKNLKEKKE SVFEYDLIKD KRTEDEKFFF KCFITINPKS
SGANKPNGBI
901 NLLKREKAND VHLISIDRGE PHLAYYTLVD KGNLIKQDT FNIIGNDPMK
TNVHDKLAAI
961 EKORDSARKD WKMINNIEM KEGYLSQVVM EIAKLVIEYN AIVVFEDLNF
GPKKGRPEVE
1021 KQVTQKLEKM LIEKLVYLVF KNEPDKXTGG VILAYQLTAP PETFKKNGKQ
TGIIYYVPAG
1081 FTSKICPVYTG FVNQLYPKYE SVSKSQEFES KFDKICYNLD KGYFEFSDY
KNFGDKAARG
1141 KWTIASFGSR LINFRNGDEN HNWDTREVYP TRELEKLLKD YSIEYGHGEC
IKAAICGESD
1201 KKKFARLTFV INTILQMENS KTGTELDYLI SPVADANGNF EDSRQAQKNN
PQKADANGAY
1261 HEGIKGIMLL GRENNQEGN KNLVIRNEE YFEFVQMPNN (SEQ ID NO:4)

Las variantes de FnCpfl aquí descritas pueden incluir la secuencia de aminoácidos de SEQ ID NO:4, con mutaciones (es decir, sustitución del aminoácido nativo por un aminoácido diferente, por ej., alanina, glicina o serina), en una o más de las siguientes posiciones: K180, E184, N607, K613, D616, N617, y/o K671 de la SEQ ID NO:4. En algunas realizaciones que no forman parte de la invención reivindicada, las variantes de FnCpfl son al menos un 80%, por ej., al menos un 85%, 90% o 95% idénticas a la secuencia de aminoácidos de SEQ ID NO:4, por ej., tienen diferencias de hasta un 5%, 10%, 15% o 20% de los residuos de SEQ ID NO:4 sustituidos, por ej., con mutaciones conservadoras, además de las mutaciones aquí descritas. En realizaciones preferidas que no forman parte de la invención reivindicada, la variante retiene la actividad deseada del progenitor, por ej., la actividad nucleasa (excepto cuando el progenitor es una nicasa o una Cpfl muerta), y/o la capacidad de interactuar con un ARN guía y ADN diana.

Para determinar el porcentaje de identidad de dos secuencias de ácido nucleico, las secuencias se alinean con fines de comparación óptima (por ej., pueden introducirse espacios en una o ambas secuencias de aminoácidos o ácidos nucleicos para una alineación óptima y las secuencias no homólogas pueden descartarse a efectos de comparación). La longitud de una secuencia de referencia alineada con fines de comparación es de al menos el 80% de la longitud de la secuencia de referencia, y en algunas realizaciones es de al menos el 90% o el 100%. Después se comparan los nucleótidos en las posiciones de aminoácidos o nucleótidos correspondientes. Cuando una posición de la primera secuencia está ocupada por el mismo nucleótido que la posición correspondiente de la segunda secuencia, las moléculas son idénticas en esa posición (tal como se utiliza aquí, "identidad" de ácido nucleico es equivalente a "homología" del ácido nucleico). El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de espacios y la longitud de cada espacio, que deben introducirse para una alineación óptima de las dos secuencias.

El porcentaje de identidad entre dos polipéptidos o secuencias de ácidos nucleicos se determina de diversas formas que se encuadran dentro de la técnica, por ejemplo, utilizando software informático disponible públicamente como Smith Waterman Alignment (Smith, T. F. y M. S. Waterman (1981) J Mol Biol 147:195-7); "BestFit" (Smith y Waterman, Advances in Applied Mathematics, 482-489 (1981)) tal como se incorpora en GeneMatcher Plus™, Schwarz y Dayhof (1979) Atlas of Protein Sequence and Structure, Dayhof, M.O., Ed, pp 353-358; programa BLAST (Basic Local Alignment Search Tool; (Altschul, S. F., W. Gish, et al. (1990) J Mol Biol 215: 403-10), BLAST-2, BLAST-P, BLAST-N, BLAST-X, WU-BLAST-2, ALIGN, ALIGN-2, CLUSTAL, o software Megalign (DNASTAR). Además, aquellos expertos en la materia pueden determinar los parámetros adecuados para medir la alineación, incluyendo los algoritmos necesarios para lograr la máxima alineación a lo largo de las secuencias comparadas. En general, para proteínas o ácidos nucleicos, la longitud de comparación puede ser cualquiera, hasta e incluyendo la longitud completa (por ej., 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% o 100%). A efectos de las presentes composiciones y métodos, se alinea al menos el 80% de la longitud total de la secuencia.

A efectos de la presente invención, la comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias puede realizarse utilizando una matriz de puntuación Blossum 62 con una penalización de espacio de 12, una penalización de longitud de espacio de 4 y una penalización de espacio del marco de lectura de 5.

Las sustituciones conservadoras incluyen típicamente sustituciones dentro de los siguientes grupos: glicina, alanina; valina, isoleucina, leucina; ácido aspártico, ácido glutámico, asparagina, glutamina; serina, treonina; lisina, arginina; y fenilalanina, tirosina.

Las variantes de Cpfl de la invención se definen en las reivindicaciones. En algunas realizaciones, los mutantes tienen alanina en lugar del aminoácido de tipo salvaje. En algunas realizaciones, los mutantes tienen cualquier aminoácido que no sea arginina o lisina (o el aminoácido nativo).

En algunas realizaciones, las variantes de Cpfl también incluyen una de las siguientes mutaciones presentadas en la Tabla A, que reducen o destruyen la actividad de la nucleasa de la Cpfl:

Tabla A

Residuos implicados en la catálisis del ADN y ARN				
	AsCpf1	LbCpf1 (+18)	LbCpf1	FnCpf1
DNA targeting (selección del ADN)	D908	D850	D832	D917
	E911	E853	E835	E920
	N913	N855	N837	H922
	Y916	Y858	Y840	Y925
	E993	E943	E925	E1006
	R1226	R1156	R1138	R1218
	S1228	S1158	S1140	S1220
	D1235	D1166	D1148	D1227
	D1263	D1198	D1180	D1255
Procesamiento de ARN	H800	H777	H759	H843
	K809	K786	K768	K852
	K860	K803	K785	K869
	F864	F807	F789	F873
Mutaciones que convierten Cpf1 en una nicasa				
	R1226A	R1156A	R1138A	R1218A

Ver, por ej., Yamano et al., Cell. 2016 May 5;165(4):949-62; Fonfara et al., Nature. 2016 Abr 28;532(7600):517-21; Dong et al., Nature. 2016 Abr 28;532(7600):522-6; y Zetsche et al., Cell. 2015 Oct 22;163(3):759-71. Hay que tener en cuenta que "LbCpf1 (+18)" se refiere a la secuencia completa de aminoácidos 1-1246 de SEQ ID NO:3, mientras que el LbCpf1 se refiere a la secuencia de LbCpf1 de Zetsche et al., también mostrada aquí como aminoácidos 1-1228 de SEQ ID NO: 11 y aminoácidos 19-1246 de SEQ ID NO:3.

Así, en algunas realizaciones, para AsCpf1, se realizan mutaciones destructoras de la actividad catalítica en D908 y E993, por ej., D908A y E993A; y para LbCpf1 mutaciones destructoras de la actividad catalítica en D832 y E925, por ej., D832A y E925A.

En algunas realizaciones, las variantes de Cpf1 también incluyen mutaciones que aumentan la especificidad (es decir, inducen sustancialmente menos efectos *off target*), por ej., como se describe en WO2018/022634. Por ejemplo, las proteínas de variantes de LbCpf1 pueden incluir una o más mutaciones en una, dos, tres, cuatro, cinco, seis o las siete posiciones siguientes: S202, N274, N278, K290, K367, K532, K609, K915, Q962, K963, K966, K1002 y/o S1003, por ej., S202A, N274A, N278A, K290A, K367A, K532A, K609A, K915A, Q962A, K963A, K966A, K1002A y/o S1003A. Las proteínas de las variantes de AsCpf1 pueden incluir una o más mutaciones en una, dos, tres, cuatro, cinco o seis de las siguientes posiciones: N178, N278, N282, R301, T315, S376, N515, K523, K524, K603, K965, Q1013, y/o K1054, por ej., N178A, N278A, N282A, R301A, T315A, S376A, N515A, K523A, K524A, K603A, K965A, Q1013A, y/o K1054A. En algunas realizaciones, las variantes de AsCpf1 comprenden mutaciones en N282A, T315A, N515A o N278A.

También se proporcionan ácidos nucleicos aislados que codifican las variantes de Cpf1, vectores que comprenden los ácidos nucleicos aislados, opcionalmente enlazados de forma operable a uno o más dominios reguladores para expresar las proteínas variantes, y células huésped, por ej., células huésped de mamífero, que comprenden los ácidos nucleicos, y opcionalmente expresan las proteínas variantes.

Las variantes aquí descritas pueden utilizarse para alterar el genoma de una célula; los métodos generalmente incluyen la expresión de las proteínas variantes en las células, junto con un ARN guía que tiene una región complementaria a una sección seleccionada del genoma de la célula. Los métodos para alterar selectivamente el genoma de una célula son conocidos en la técnica, ver, por ej., US 8.993.233; US 20140186958; US 9.023.649; WO/2014/099744; WO 2014/089290; WO2014/144592; WO144288;

WO2014/204578; WO2014/152432; WO2115/099850; US8,697,359; US20160024529; US20160024524; US20160024523; US20160024510; US20160017366; US20160017301; US20150376652; US20150356239;

- US20150315576; US20150291965; US20150252358; US20150247150; US20150232883; US20150232882;
 US20150203872; US20150191744; US20150184139; US20150176064; US20150167000; US20150166969;
 US20150159175; US20150159174; US20150093473; US20150079681; US20150067922; US20150056629;
 US20150044772; US20150024500; US20150024499; US20150020223; US20140356867; US20140295557;
 5 US20140273235; US20140273226; US20140273037; US20140189896; US20140113376; US20140093941;
 US20130330778; US20130288251; US20120088676; US20110300538; US20110236530; US20110217739;
 US20110002889; US20100076057; US20110189776; US20110223638; US20130130248; US20150050699;
 US20150071899; US20150045546; US20150031134; US20150024500; US20140377868; US20140357530;
 US20140349400; US20140335620; US20140335063; US20140315985; US20140310830; US20140310828;
 10 US20140309487; US20140304853; US20140298547; US20140295556; US20140294773; US20140287938;
 US20140273234; US20140273232; US20140273231; US20140273230; US20140271987; US20140256046;
 US20140248702; US20140242702; US20140242700; US20140242699; US20140242664; US20140234972;
 US20140227787; US20140212869; US20140201857; US20140199767; US20140189896; US20140186958;
 US20140186919; US20140186843; US20140179770; US20140179006; US20140170753; WO/2008/108989;
 15 WO/2010/054108; WO/2012/164565 WO/2013/098244; WO/2013/176772; Makarova et al., "Evolution and
 classification of the CRISPR-Cas systems" 9(6) Nature Reviews Microbiology 467-477 (1-23) (Jun. 2011);
 Wiedenheft et al., "RNA-guided genetic silencing systems in bacteria and archaea" 482 Nature 331-338 (16 feb.
 2012); Gasiunas et al., "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive
 immunity in bacteria" 109(39) Proceedings of the National Academy of Sciences USA E2579-E2586 (4 sept. 2012);
 20 Jinek et al., "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity" 337 Science
 816-821 (17 ago. 2012); Carroll, "A CRISPR Approach to Gene Targeting" 20(9) Molecular Therapy 1658-1660 (sep.
 2012); U.S. Appl. No. 61/652,086, presentado el 25 de mayo de 2012; Al-Attar et al., Clustered Regularly Interspaced
 Short Palindromic Repeats (CRISPRs): The Hallmark of an Ingenious Antiviral Defense Mechanism in Prokaryotes,
 Biol Chem. (2011) vol. 392, Edición 4, pp. 277-289; Hale et al., Essential Features and Rational Design of CRISPR
 25 RNAs That Function With the Cas RAMP Module Complex to Cleave RNAs, Molecular Cell, (2012) vol. 45, Edición
 3, 292-302.

Las proteínas variantes aquí descritas pueden utilizarse en lugar o además de cualquiera de las proteínas Cas9 o
 CpfI descritas en las referencias anteriores, o en combinación con mutaciones análogas descritas en las mismas,
 con un ARN guía apropiado para la CpfI seleccionada, es decir, con ARN guía que se dirija a secuencias distintas
 30 de PAM de tipo salvaje, por ej., que tenga secuencias PAM según la siguiente Tabla B.

TABLA B		
Proteína variante	PAM más fuerte	PAM más débil
AsCpfI S170R	TTTN, CTTN, GTTN	TATN, TCTN, TTAN, TTCN, TTGN
AsCpfI E174R	TTTN, CTTN, GTTN, TTCN	TATN, TCTN, TTAN, TTGN
AsCpfI S542K	TTTN, GTTN, TTCN	CTTN, TCTN, TTAN
AsCpfI S542Q	TTTN	TTCN
AsCpfI S542R	TTTN, ATTN, CTTN, GTTN, TCTN, TTCN, ATCN, CCCN, CCTN, GCTN, GGTN, TCCN	TGTN, TATN, CTCN, TGCN
AsCpfI N551R	TTTN	GTTN
AsCpfI N552R		TTTN
AsCpfI T167A/T539K	TTTN, GTTN	
AsCpfI T167A/T539R	TTTN, GTTN	
AsCpfI E174R/S542R	TTTN, CTTN, TTCN	
AsCpfI S542R/K548R	TTTN, CTTN, GTTN, TTCN	TATN, TGTN, TTAN
AsCpfI S542R/N551R	TTTN, GTTN, TTCN	CTTN
AsCpfI S542R/N552R	TTTN	GTTN

ES 2 992 912 T3

AsCpfl K548R/N551R	TTTN, GTTN	
AsCpfl K548R/N552R		TTTN
AsCpfl S542R/M604A	TTTN	TTCN
AsCpfl S542R/K607H	TTTN, TTCN	
AsCpfl E174R/S542R/K548R	TTTN, CTTN, GTTN, ATTN, TATN, TGTN, TTAN, TTCN, AGTN, CATN, CCCN, CGTN, CTCN, GATN, GCTN, GGTN, GTCN, TACN, TCCN, TGCN, ACCN	TCTN, TTGN, AAAN, ACTN, ATCN, CCTN, CTAN
AsCpfl S542R/K548R/N551R	TTTN, GTTN, TTCN	CTTN, TATN, TGTN, TTAN
AsCpfl S170R/S542R/K607R	TTTN, TTCN, TCCN, TCTN, ACCN	GTTN, TTAN
AsCpfl E174R/S542R/K607H	CTTN, TCTN, TTCN, TCCN, TTTN	CCCN, ACCN
AsCpfl E174R/S542R/K607R	TTTN, TTCN, TCCN, CCCN, ACCN, GCCN	CTTN, GTTN, TCTN, TTAN, TTGN
AsCpfl E174R/S542R/K548R/N551R	TTTN, CTTN, GTTN, TTCN, TCCN, CCCN, ACCN	TATN
AsCpfl E174R/S542R/K548R/N552R	TTTN, CTTN, GTTN	TATN
AsCpfl E174R/S542R/K548V/N552R	TTTN, CTTN, GTTN, TATN	TCTN, TGTN, TTCN, TCCN
AsCpfl S170R/S542R/K548V/N552R	TTTN, GTTN, TATN,	CTTN, TGTN, TTCN
LbCpfl T152R	TTTN, TTCN	
LbCpfl T152K	TTTN, TTCN	

(Continúa)

TABLA B		
Proteína variante	PAM más fuerte	PAM más débil
LbCpfl D156R	TTTN, TTCN	
LbCpfl D156K	TTTN, TTCN	
LbCpfl G532R	TTTN, TTCN	
LbCpfl K538R	TTTN	TTCN
LbCpfl D156R/G532R/K538R	TTTN, CTTN, GTTN, TTAN, TTCN, TTGN, TCCN	TATN, TCTN
FnCpfl K180R	TTTN, CTTN, GTTN, NTTN, TCTN	TTAN, TTCN
FnCpfl N607R	TTTN, CTTN, GTTN, NTTN, TCTN, TTCN	TTAN
FnCpfl K613R	TTTN, CTTN, GTTN, NTTN,	TTCN, TGTN
FnCpfl K613V	TTTN, CTTN, GTTN, NTTN,	TATN, TGTN, TTCN

FnCpfl D616R	TTTN, CTTN, GTTN, NTTN, TCTN, TTCN	TTAN
FnCpfl N617R	TTTN, CTTN, GTTN, NTTN, TCTN	TTCN
FnCpfl K671R	TTTN, TCTN	CTTN, GTTN, NTTN, TTCN
FnCpfl K671H	TTTN, CTTN, GTTN, TCTN, NTTN	TTCN
FnCpfl K607R/K613V	TTTN, CTTN, GTTN, NTTN, TGTN, TTCN	TATN, TCTN, TTAN
FnCpfl K607R/K613V/D616R	TTTN, GTTN, GTTN, NTTN, TATN, TCTN, TGTN, TTAN, TTCN	
FnCpfl K607R/K613R/D616R	TTTN, GTTN, GTTN, NTTN, TGTN, TTAN, TTCN	TCTN,

Las variantes aquí descritas también pueden utilizarse en métodos de detección de un ADNss o ADNds diana en una muestra *in vitro*, por ej., como se describe en US20170362644; East-Seletsky et al., Nature. 2016 Oct 13; 538(7624): 270-273; Gootenberg et al., Science. 2017 Apr 28; 356(6336): 438-442; Gootenberg et al., Science 10.1126/science.aag0179 (2018); Chen et al., Science. 2018 Feb 15. Pii: eaar6245; Science. 2018 Feb 15. Pii: eaaq0179; y WO2017219027A1. En estos métodos, la unión de la variante a su diana induce una actividad Dnasa no específica contra otras dianas. Los métodos incluyen el contacto de una muestra de la que se sabe o se sospecha que incluye un ADNss o ADNds diana con la proteína de fusión (o una pluralidad de proteínas de fusión), ARN guía cognado que funciona con esas proteínas de fusión y ADN detector marcado (por ej., un ADNss reportador de 3-30 nts, 3-20, 5-20, 5-15 u otra longitud adecuada). Cuando una proteína de fusión se une a su diana, la actividad de ADN no específica escinde los ADN detectores, produciendo una señal. Los métodos para medir la señal del ADN detector marcado son conocidos en la técnica y pueden incluir, por ejemplo, la detección de una o más nanopartículas de oro, un fluoróforo, polarización de fluorescencia, transición/dispersión de fase coloidal, señales electroquímicas y señales basadas en semiconductores. En algunas realizaciones, el ADN detector marcado produce una cantidad de señal detectable antes de escindirse, y la cantidad de señal detectable se reduce cuando el ADN detector marcado se escinde. Alternativamente, el ADN detector marcado puede producir una primera señal detectable antes de ser escindido y una segunda señal detectable cuando el ADN detector marcado se escinde. En algunas realizaciones, el ADN detector marcado comprende un par quencher/fluor. En algunas realizaciones, también se incluye Csm6, una enzima auxiliar asociada a CRISPR.

Además, las variantes aquí descritas, pueden utilizarse en proteínas de fusión en lugar de la Cas9 de tipo salvaje u otras mutaciones de Cas9 (como la dCas9 o la Cas9 nicasa) como se conoce en la técnica, por ej., una proteína de fusión con un dominio funcional heterólogo como se describe en US 8,993,233; US 20140186958; US 9,023,649; WO/2014/099744; WO 2014/089290; WO2014/144592; WO144288; WO2014/204578; WO2014/152432; WO2115/099850; US8,697,359; US2010/0076057; US2011/0189776; US2011/0223638; US2013/0130248; WO/2008/108989; WO/2010/054108; WO/2012/164565; WO/2013/098244; WO/2013/176772; US20150050699; US 20150071899 y WO 2014/124284. Por ejemplo, las variantes, que comprenden preferentemente una o más mutaciones reductoras o destructoras de nucleasas, pueden fusionarse en el extremo N o C del Cpfl a un dominio de activación transcripcional (por ej., un dominio de activación transcripcional del dominio VP16 del virus del herpes simple (Sadowski et al., 1988, Nature, 335:563-564) o VP64; el dominio p65 del factor de transcripción celular NF-kappaB (Ruben et al., 1991, Science, 251:1490-93); o un efector tripartito fusionado a dCas9, compuesto de activadores VP64, p65 y Rta (VPR) enlazados en tándem, Chavez et al., Nat Methods. 2015 Abr;12(4):326-8) u otros dominios funcionales heterólogos (por ej., represores transcripcionales (por ej., KRAB, ERD, SID, y otros, por ej., aminoácidos 473-530 del dominio represor (ERD) del factor represor ets2 (ERF), aminoácidos 1-97 del dominio KRAB de KOX1, o aminoácidos 1-36 del dominio de interacción Mad mSIN3 (SID); ver Beerli et al., PNAS USA 95:14628-14633 (1998)) o silenciadores como la Proteína 1 Heterocromatina (HP1, también conocida como swi6), por ej., HP1 α o HP1 β ; proteínas o péptidos que podrían reclutar ARN no codificantes largos (ARNlnc) fusionados a una secuencia fija de unión a ARN, como las unidas por la proteína de cubierta MS2, la endorribonucleasa Csy4 o la proteína lambda N; editores de bases (enzimas que modifican el estado de metilación del ADN (por ej., ADN metiltransferasa (DNMT) o proteínas TET); o enzimas que modifican subunidades de histonas (por ej., histona acetiltransferasas (HAT), histona desacetilasas (HDAC), histona metiltransferasas (por ej., para la metilación de residuos de lisina o arginina) o histona desmetilasas (por ej., para la desmetilación de residuos de lisina o arginina)) como se conocen en la técnica. En la técnica se conocen diversas secuencias para dichos dominios, por ej., un dominio que cataliza la hidroxilación de citosinas metiladas en ADN. Las proteínas ejemplares incluyen la familia Ten-Eleven-Translocation (TET)1-3, enzimas que convierten la 5-metilcitosina (5-mC) en 5-hidroximetilcitosina (5-hmC) en ADN.

Las secuencias de TET1-3 humana son conocidas en la técnica y se muestran en la tabla siguiente:

Gen	Números de acceso GenBank	
	Aminoácido	Ácido nucleico
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1 (var 1) NP_060098.3 (var 2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1
* La variante (1) representa la transcripción más larga y codifica la isoforma más larga (a). La variante (2) difiere en secuencia codificante de 5' UTR y de 3' UTR comparado con la variante 1. La isoforma resultante (b) es más corta y tiene una terminación C distinta en comparación con la isoforma a.		

En algunas realizaciones, puede incluirse toda o parte de la secuencia de longitud completa del dominio catalítico, por ej., un módulo catalítico que comprende la extensión rica en cisteína y el dominio 2OGFeDO codificado por 7 exones altamente conservados, por ej., el dominio catalítico Tet1 que comprende los aminoácidos 1580-2052, Tet2 que comprende los aminoácidos 1290-1905 y Tet3 que comprende los aminoácidos 966-1678. Ver, por ej., la Fig. 1 de Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009, 27 Jun, para un alineamiento que ilustra los residuos catalíticos clave en las tres proteínas Tet, y los materiales suplementarios del mismo (disponibles en el sitio ftp://ftp.ncbi.nih.gov/pub/34ravind/DONS/supplementary_material_DONS.html) para secuencias de longitud completa (ver, por ej., seq 2c); en algunas realizaciones, la secuencia incluye los aminoácidos 1418-2136 de Tet1 o la región correspondiente en Tet2/3.

Otros módulos catalíticos pueden proceder de las proteínas identificadas en Iyer et al., 2009.

En algunas realizaciones, el dominio funcional heterólogo es un editor de bases, por ej., una deaminasa que modifica bases de ADN de citosina, por ej., una citidina deaminasa de la familia de las deaminasas similares al polipéptido catalítico, enzima editora de ARNm de la apolipoproteína B (APOBEC), incluyendo APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/E, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 (ver, por ej., Yang et al., J Genet Genomics. 2017, 20 Sep;44(9):423-437); citidina deaminasa inducida por activación (AID), por ej., citidina deaminasa inducida por activación (AIC-DA), citosina deaminasa 1 (CDA1) y CDA2, y citosina deaminasa que actúa sobre ARNt (CDAT). La siguiente tabla proporciona secuencias de ejemplo; también pueden usarse otras secuencias.

Deaminasa	Números de acceso GenBank	
	Ácido nucleico	Aminoácido
hAID/AICDA	NM_020661.3 isoforma 1 NM_020661.3 isoforma 2	NP_065712.1 variante 1 NP_065712.1 variante 2
APOBEC1	NM_001644.4 isoforma a NM_005889.3 isoforma b	NP_001635.2 variante 1 NP_005880.2 variante 3
APOBEC2	NM_006789.3	NP_006780.1
APOBEC3A	NM_145699.3 isoforma a NM_001270406.1 isoforma b	NP_663745.1 variante 1 NP_001257335.1 variante 2
APOBEC3B	NM_004900.4 isoforma a	NP_004891.4 variante 1

(continúa)

	Números de acceso GenBank	
	NM_001270411.1 isoforma b	NP_001257340.1 variante 2
APOBEC3C	NM_014508.2	NP_055323.2

APOBEC3D/E	NM_152426.3	NP_689639.2
APOBEC3F	NM_145298.5 isoforma a NM_001006666.1 isoforma b	NP_660341.2 variante 1 NP_001006667.1 variante 2
APOBEC3G	NM_021822.3 (isoforma a)	NP_068594.1 (variante 1)
APOBEC3H	NM_001166003.2	NP_001159475.2 (variante SV-200)
APOBEC4	NM_203454.2	NP_982279.1
CDA1*	NM_127515.4	NP_179547.1
* de <i>Saccharomyces cerevisiae</i> S288C		

En algunas realizaciones, el dominio funcional heterólogo es una deaminasa que modifica bases de ADN de adenosina, por ej., la deaminasa es una adenosina deaminasa 1 (ADA1), ADA2; adenosina deaminasa que actúa sobre ARN 1 (ADAR1), ADAR2, ADAR3 (ver, por ej., Savva et al., Genome Biol. 2012 28 Dic;13(12):252); adenosina deaminasa que actúa sobre ARNt 1 (ADAT1), ADAT2, ADAT3 (ver Keegan et al., RNA. 2017 Sep;23(9):1317-1328 y Schaub and Keller, Biochimie. 2002 Ag;84(8):791-803); y adenosina deaminasa específica de ARNt natural o modificada por ingeniería (TadA) (ver, por ej., Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471) (NP_417054.2 (*Escherichia coli* str. K-12 substr. MG1655); ver, por ej., Wolf et al., EMBO J. 2002 Jul 15;21(14):3841-51). La siguiente tabla proporciona secuencias de ejemplo; también pueden usarse otras secuencias.

	Números de acceso GenBank	
Deaminasa	Ácido nucleico	Aminoácido
ADA (ADA1)	NM_000022.3 variante 1	NP_000013.2 isoforma 1
ADA2	NM_001282225.1	NP_001269154.1
ADAR	NM_001111.4	NP_001102.2
ADAR2 (ADARB1)	NM_001112.3 variante 1	NP_001103.1 isoforma 1
ADAR3 (ADARB2)	NM_018702.3	NP_061172.1
ADAT1	NM_012091.4 variante 1	NP_036223.2 isoforma 1
ADAT2	NM_182503.2 variante 1	NP_872309.2 isoforma 1
ADAT3	NM_138422.3 variante 1	NP_612431.2 isoforma 1

En algunas realizaciones, el dominio funcional heterólogo es una enzima, dominio o péptido que inhibe o mejora las vías endógenas de reparación del ADN o de reparación por escisión de bases (BER), por ej., timina ADN glicosilasa (TDG; GenBank N° acceso NM_003211.4 (ácido nucleico) y NP_003202.3 (proteína)) o uracilo ADN glicosilasa (UDG, también conocida como uracilo N-glicosilasa, o UNG); GenBank N.° acceso NM_003362.3 (ácido nucleico) y NP_003353.1 (proteína)) o inhibidor de uracilo ADN glicosilasa (UGI) que inhibe la escisión de uracilo mediada por UNG para iniciar BER. (ver, por ej., Mol et al., Cell 82, 701-708 (1995); Komor et al., Nature. 2016 May 19;533(7603)); o proteínas de unión a extremos del ADN como Gam, que es una proteína del bacteriófago Mu que se une a extremos libres del ADN, inhibiendo las enzimas de reparación del ADN y dando lugar a una edición más precisa (menos ediciones de bases no intencionadas; Komor et al., Sci Adv. 2017 Aug 30;3(8):eaao4774).

En algunas realizaciones, puede utilizarse toda o parte de la proteína, por ej., al menos un dominio catalítico que conserve la función prevista de la enzima.

En algunas realizaciones, el dominio funcional heterólogo es una biomolécula, y comprende todo o parte de (por ej., el dominio de unión al ADN de) la proteína de cubierta MS2, la endorribonucleasa Csy4, o la proteína lambda N. Estas proteínas pueden utilizarse para reclutar moléculas de ARN que contienen una estructura de bucle de tallo específica a una localización especificada por las secuencias diana de ARNg dCpf1. Por ejemplo, una variante de dCpf1 fusionada con la proteína de cubierta MS2, la endorribonucleasa Csy4 o lambda N puede utilizarse para

reclutar un ARN no codificante largo (ARNlnc) como XIST o HOTAIR; ver, por ej., Keryer-Bibens et al., Biol. Cell 100:125-138 (2008), que está enlazado a la secuencia de unión Csy4, MS2 o lambda N. Alternativamente, la secuencia de unión de las proteínas Csy4, MS2 o lambda N puede enlazarse con otra proteína, por ej., como se describe en Keryer-Bibens et al., supra, y la proteína puede dirigirse al sitio de unión de la variante dCpf1 utilizando los métodos y composiciones aquí descritos. En algunas realizaciones, el Csy4 es catalíticamente inactivo. En algunas realizaciones, la variante Cpf1, preferentemente una variante dCpf1, se fusiona a FokI como se describe en US 8,993,233; US 20140186958; US 9,023,649; WO/2014/099744; WO 2014/089290; WO2014/144592; WO144288; WO2014/204578; WO2014/152432; WO2115/099850; US8,697,359; US2010/0076057; US2011/0189776; US2011/0223638; US2013/0130248; WO/2008/108989; WO/2010/054108; WO/2012/164565; WO/2013/098244; WO/2013/176772; US20150050699; US 20150071899 y WO 2014/204578.

En algunas realizaciones, las proteínas de fusión incluyen un enlazador entre la variante Cpf1 y los dominios funcionales heterólogos. Los enlazadores que pueden utilizarse en estas proteínas de fusión (o entre proteínas de fusión en una estructura concatenada) pueden incluir cualquier secuencia que no interfiera con la función de las proteínas de fusión. En realizaciones preferidas, los enlazadores son cortos, por ej., 2-20 aminoácidos, y son típicamente flexibles (es decir, comprenden aminoácidos con un alto grado de libertad como glicina, alanina y serina). En algunas realizaciones, el enlazador comprende una o más unidades consistentes en GGGS (SEQ ID NO:12) o GGGGS (SEQ ID NO:13), por ej., dos, tres, cuatro o más repeticiones de la unidad GGGS (SEQ ID NO:12) o GGGGS (SEQ ID NO:13). También se pueden usar otras secuencias de enlace.

En algunas realizaciones, la proteína variante incluye una secuencia peptídica penetrante celular que facilita la entrega al espacio intracelular, por ej., péptido TAT derivado de HIV, penetratinas, transportanos o péptidos penetrantes celulares derivados de hCT, ver, por ej., Caron et al., (2001) Mol Ther. 3(3):310-8; Langel, Cell-Penetrating Peptides: Processes and Applications (CRC Press, Boca Raton FL 2002); El-Andaloussi et al., (2005) Curr Pharm Des. 11(28):3597-611; y Deshayes et al., (2005) Cell Mol Life Sci. 62(16):1839-49.

Los péptidos penetrantes celulares (CPP) son péptidos cortos que facilitan el movimiento de una amplia gama de biomoléculas a través de la membrana celular hacia el citoplasma u otros orgánulos, por ej. las mitocondrias y el núcleo. Entre los ejemplos de moléculas que pueden transportarse mediante CPP se incluyen fármacos terapéuticos, ADN plasmídico, oligonucleótidos, ARNs, ácido peptidínucleico (PNA), proteínas, péptidos, nanopartículas y liposomas. Los CPP suelen tener 30 aminoácidos o menos, se derivan de proteínas naturales o no naturales o de secuencias químicas, y contienen una gran abundancia relativa de aminoácidos cargados positivamente, por ej. lisina o arginina, o un patrón alterno de aminoácidos polares y no polares. Entre los CPP comúnmente utilizados en la técnica se incluyen Tat (Frankel et al., (1988) Cell. 55:1189-1193, Vives et al., (1997) J. Biol. Chem. 272:16010-16017), penetratina (Derossi et al., (1994) J. Biol. Chem. 269:10444-10450), polyarginine peptide sequences (Wender et al., (2000) Proc. Natl. Acad. Sci. USA 97:13003-13008, Futaki et al., (2001) J. Biol. Chem. 276:5836-5840), and transportan (Pooga et al., (1998) Nat. Biotechnol. 16:857-861).

Los CPP pueden enlazarse con su carga mediante estrategias covalentes o no covalentes. Se conocen métodos para la unión covalente de un CPP y su carga, por ej., el *cross-linking* químico (unión química mediante un enlace covalente) (Stetsenko et al., (2000) J. Org. Chem. 65:4900-4909, Gait et al. (2003) Cell. Mol. Life. Sci. 60:844-853) o la clonación de una proteína de fusión (Nagahara et al., (1998) Nat. Med. 4:1449-1453). El acoplamiento no covalente entre la carga y los CPP anfipáticos cortos que comprenden dominios polares y no polares se establece a través de interacciones electrostáticas e hidrofóbicas.

Los CPP se han utilizado en la técnica para introducir biomoléculas potencialmente terapéuticas en las células. Algunos ejemplos son la ciclosporina enlazada a la poliarginina para la inmunosupresión (Rothbard et al., (2000) Nature Medicine 6(11):1253-1257), ARNs contra ciclina B1 enlazado a un CPP llamado MPG para inhibir la tumorigénesis (Crombez et al., (2007) Biochem Soc. Trans. 35:44-46), péptidos supresores tumorales p53 enlazados a CPP para reducir el crecimiento de células cancerosas (Takenobu et al., (2002) Mol. Cancer Ther. 1(12):1043-1049, Snyder et al., (2004) PloS Biol. 2:E36), y formas dominantes negativas de Ras o fosfoinositol 3 quinasa (PI3K) fusionadas a Tat para tratar el asma (Myou et al., (2003) J. Immunol. 171:4399-4405).

Los CPP se han utilizado para transportar agentes de contraste al interior de las células para la obtención de imágenes y aplicaciones de biodetección. Por ejemplo, la proteína verde fluorescente (GFP) unida a Tat se ha utilizado para marcar células cancerosas (Shokolenko et al., (2005) DNA Repair 4(4):511-518). Tat conjugado con puntos cuánticos se ha utilizado para atravesar con éxito la barrera hematoencefálica para la visualización del cerebro de rata (Santra et al., (2005) Chem. Commun. 3144-3146). Los CPP también se han combinado con técnicas de resonancia magnética para la obtención de imágenes de células (Liu et al., (2006) Biochem. And Biophys. Res. Comm. 347(1):133-140). Ver también Ramsey and Flynn, Pharmacol Ther. 2015 Jul 22. Pii: S0163-7258(15)00141-2.

En algunas realizaciones, alternativa o adicionalmente, las proteínas variantes pueden incluir una secuencia de localización nuclear, por ej., el antígeno T grande de SV40 NLS (PKKKRRV (SEQ ID NO:16)) y nucleoplasmina NLS (KRPAATKKAGQAKKKK (SEQ ID NO:7)). Otros NLS son conocidos en la técnica; ver, por ej., Cokol et al., EMBO Rep. 2000 Nov 15; 1(5): 411-415; Freitas and Cunha, Curr Genomics. 2009 Dic; 10(8): 550-557.

- 5 En algunas realizaciones, las variantes incluyen una fracción que tiene una alta afinidad por un ligando, por ejemplo GST, FLAG o secuencias de hexahistidina. Tales marcadores de afinidad pueden facilitar la purificación de proteínas variantes recombinantes.

Para los métodos en los que las proteínas variantes se administran a las células, las proteínas pueden producirse utilizando cualquier método conocido en la técnica, por ej., por traducción *in vitro*, o expresión en una célula huésped adecuada a partir de ácido nucleico que codifica la proteína variante; se conocen varios métodos en la técnica para producir proteínas. Por ejemplo, las proteínas pueden producirse y purificarse a partir de levaduras, *E. coli*, líneas celulares de insectos, plantas, animales transgénicos o células cultivadas de mamíferos; ver, por ej., Palomares et al., "Production of Recombinant Proteins: Challenges and Solutions," Methods Mol Biol. 2004;267:15-52.

Además, las proteínas variantes pueden estar enlazadas a una fracción que facilite su transferencia al interior de una célula, por ej., una nanopartícula lipídica, opcionalmente con un enlazador que se escinde una vez que la proteína está dentro de la célula. Ver, por ej., LaFountaine et al., Int J Pharm. 2015 Ago 13;494(1): 180-194.

Sistemas de expresión

Para utilizar las variantes de Cpfl aquí descritas, puede ser deseable expresarlas a partir de un ácido nucleico que las codifique. Esto puede realizarse de varias maneras. Por ejemplo, el ácido nucleico que codifica la variante de Cpfl puede clonarse en un vector intermedio para la transformación en células procariotas o eucariotas para la replicación y/o expresión. Los vectores intermedios son típicamente vectores procariotas, por ej., plásmidos, o vectores lanzadera, o vectores de insectos, para el almacenamiento o manipulación del ácido nucleico que codifica la variante Cpfl para la producción de la variante Cpfl. El ácido nucleico que codifica la variante Cpfl también puede clonarse en un vector de expresión, para su administración a una célula vegetal, una célula animal, preferiblemente una célula de mamífero o una célula humana, una célula fúngica, una célula bacteriana o una célula protozoaria.

Para obtener la expresión, una secuencia que codifica una variante de Cpfl suele subclonarse en un vector de expresión que contiene un promotor para dirigir la transcripción. Los promotores bacterianos y eucarióticos adecuados son bien conocidos en la técnica y se describen, por ej., en Sambrook et al., Molecular Cloning, A Laboratory Manual (3ª ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); y Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Se dispone de sistemas de expresión bacteriana para expresar la proteína modificada mediante ingeniería, por ej., en *E. coli*, *Bacillus* sp. y *Salmonella*. (Palva et al., 1983, Gene 22:229-235). Existen kits en el comercio para este tipo de sistemas de expresión. Los sistemas de expresión eucariótica para células de mamíferos, levaduras e insectos son bien conocidos en la técnica y también están disponibles comercialmente.

El promotor utilizado para dirigir la expresión de un ácido nucleico depende de la aplicación en particular. Por ejemplo, un promotor constitutivo fuerte se utiliza típicamente para la expresión y purificación de proteínas de fusión. Por el contrario, cuando la variante Cpfl se va a administrar *in vivo* para la regulación génica, se puede utilizar un promotor constitutivo o inducible, dependiendo del uso particular de la variante Cpfl. Además, un promotor preferido para la administración de la variante Cpfl puede ser un promotor débil, como HSV TK o un promotor que tenga una actividad similar. El promotor también puede incluir elementos que respondan a la transactivación, por ej., elementos de respuesta a la hipoxia, elementos de respuesta Gal4, elementos de respuesta de represor lac y sistemas de control de moléculas pequeñas como los sistemas regulados por tetraciclina y el sistema RU-486 (ver, por ej., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88: 1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

Además del promotor, el vector de expresión contiene típicamente una unidad de transcripción o casete de expresión que contiene todos los elementos adicionales necesarios para la expresión del ácido nucleico en células huésped, procariotas o eucariotas. Así pues, un casete de expresión típico contiene un promotor enlazado de forma operable, por ej., a la secuencia de ácido nucleico que codifica la variante de Cpfl, y cualquier señal necesaria, por ej., para la poliadenilación eficiente de la transcripción, la terminación transcripcional, los sitios de unión a ribosomas o la terminación de la traducción. Los elementos adicionales del casete pueden incluir, por ej., potenciadores y señales intrónicas empalmadas heterólogas.

El vector de expresión particular utilizado para transportar la información genética a la célula se selecciona en función del uso previsto de la variante Cpfl, por ej., expresión en plantas, animales, bacterias, hongos, protozoos,

etc. Los vectores de expresión bacteriana estándar incluyen plásmidos como los plásmidos basados en pBR322, pSKF, pET23D, y sistemas de expresión de fusión de marcadores disponibles comercialmente como GST y LacZ.

Los vectores de expresión que contienen elementos reguladores de virus eucariotas se utilizan a menudo en vectores de expresión eucariotas, por ej., vectores SV40, vectores del virus del papiloma y vectores derivados del virus de Epstein-Barr. Otros vectores eucariotas de ejemplo incluyen pMSG, pAV009/A+, pMT010/A+, pMAMneo-5, baculovirus pDSVE, y cualquier otro vector que permita la expresión de proteínas bajo la dirección del promotor temprano de SV40, el promotor tardío de SV40, el promotor de la metalotioneína, el promotor del virus del tumor mamario murino, el promotor del virus del sarcoma de Rous, el promotor de la polihedrina, u otros promotores que hayan demostrado efectividad para la expresión en células eucariotas.

Los vectores para expresar las variantes de Cpfl pueden incluir promotores de ARN Pol III para dirigir la expresión de los ARN guía, por ej., los promotores H1, U6 o 7SK. Estos promotores humanos permiten la expresión de las variantes de Cpfl en células de mamífero tras la transfección del plásmido.

Algunos sistemas de expresión tienen marcadores para la selección de líneas celulares transfectadas de forma estable, como la timidina quinasa, la higromicina B fosfotransferasa y la dihidrofolato reductasa. También son adecuados los sistemas de expresión de alto rendimiento, como el uso de un vector de baculovirus en células de insecto, con la secuencia que codifica el ARNg bajo la dirección del promotor de poliedrina u otros promotores fuertes de baculovirus.

Los elementos que suelen incluirse en los vectores de expresión también incluyen un replicón que funciona en *E. coli*, un gen que codifica la resistencia a los antibióticos para permitir la selección de bacterias que albergan plásmidos recombinantes, y sitios de restricción únicos en regiones no esenciales del plásmido para permitir la inserción de secuencias recombinantes.

Los métodos estándar de transfección se utilizan para producir líneas celulares de bacterias, mamíferos, levaduras o insectos que expresan grandes cantidades de proteínas, que luego se purifican mediante técnicas estándar (ver, por ej., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). La transformación de células eucariotas y procariotas se realiza según las técnicas habituales (ver, por ej., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Puede utilizarse cualquiera de los procedimientos conocidos para introducir secuencias de nucleótidos extrañas en células huésped. Estos incluyen el uso de transfección de fosfato cálcico, polibreno, fusión de protoplastos, electroporación, nucleofección, liposomas, microinyección, ADN desnudo, vectores plasmídicos, vectores virales, tanto episomales como integradores, y cualquiera de otros métodos conocidos para introducir ADN genómico clonado, ADNc, ADN sintético u otro material genético extraño en una célula huésped (ver, por ej., Sambrook et al., supra). Sólo es necesario que el procedimiento particular de ingeniería genética utilizado sea capaz de introducir con éxito al menos un gen en la célula huésped capaz de expresar la variante de Cpfl.

La presente invención también incluye los vectores y las células que comprenden los vectores.

También se proporcionan composiciones y kits que comprenden las variantes aquí descritas, que no forman parte de la invención reivindicada. En algunas realizaciones que no forman parte de la invención reivindicada, los kits incluyen las proteínas de fusión y un ARN guía cognado (es decir, un ARN guía que se une a la proteína y la dirige a una secuencia diana apropiada para esa proteína). En algunas realizaciones que no forman parte de la invención reivindicada, los kits también incluyen ADN detector marcado, por ej., para uso en un método de detección de un ADNss o ADNds diana. Los ADN detectores marcados son conocidos en la técnica, por ej., como se describe en US20170362644; East-Seletsky et al., Nature. 2016 Oct 13; 538(7624): 270-273; Gootenberg et al., Science. 2017 Apr 28; 356(6336): 438-442, y WO2017219027A1, y pueden incluir ADN detector marcado que comprende un par de transferencia de energía por resonancia de fluorescencia (FRET) o un par quencher/fluor, o ambos. Los kits también pueden incluir uno o más reactivos adicionales, por ej., enzimas adicionales (como ARN polimerasas) y tampones, por ej., para el uso en un método aquí descrito.

EJEMPLOS

La invención se describe con más detalle en los siguientes ejemplos, que no limitan el alcance de la invención descrita en las reivindicaciones. Los ejemplos que quedan fuera del alcance de las reivindicaciones se facilitan a efectos comparativos.

Métodos

Los siguientes materiales y métodos se utilizaron en los Ejemplos siguientes, salvo que se indique lo contrario.

Plásmidos y oligonucleótidos.

Las secuencias del sitio diana para los ARNcr y las secuencias de oligonucleótidos utilizadas en los Ejemplos 1B, 4B y 5-8 están disponibles en las Tablas 2A-2D y 3A-3D respectivamente. Los plásmidos de expresión humana para AsCas12a, LbCas12a, FnCas12a y MbCas12a de tipo salvaje (SQT1659, SQT1665, AAS1472, AAS2134, respectivamente) se generaron subclonando los marcos abiertos de lectura de nucleasa de los plásmidos pY010, pY016, pY004 y pY014, respectivamente (plásmidos Addgene 69982, 69988, 69976 y 69986; donación de Feng Zhang) en los sitios NotI y AgeI de pCAG-CFP (plásmido Addgene 11179; donación de Connie Cepko). Los plásmidos de expresión de proteínas se generaron clonando los marcos abiertos de lectura optimizados para codones humanos de As, Fn y MbCas12a, y el marco abierto de lectura optimizado para codones bacterianos de LbCas12a (del plásmido Addgene 79008; una donación de Jin Soo Kim) en los sitios NcoI y FseI de pET28b-Cas9 (plásmido Addgene 47327; donación de Alex Schier) para generar BPK3541, RTW656, RTW660 y RTW645, respectivamente. Todas las variantes de Cas12a, las construcciones activadoras y las fusiones editoras de bases se generaron mediante clonación molecular estándar y ensamblaje isotérmico. Se generaron plásmidos de expresión de células humanas para los ARNcr Cas12a mediante hibridación y ligado de oligonucleótidos correspondientes a dúplexes de secuencias espaciadoras en BPK3079, BPK3082, BPK4446 y BPK4449 BsmBI-digeridos para la transcripción impulsada por el promotor U6 de los ARNcr de As, Lb, Fn y MbCas12a respectivamente. Los plásmidos de sustrato para las reacciones de escisión *in vitro* se generaron clonando los sitios diana en los sitios NheI e HindIII de pUC19. Los plásmidos para la transcripción *in vitro* de los ARNcr Cas2a se generaron por hibridación y ligado de oligonucleótidos correspondientes a dúplexes de secuencias espaciadoras en MSP3491, MSP3495, RTW763 y RTW767 BsaI-digeridos para la transcripción impulsada por el promotor T7 del ARNcr de As, Lb, Fn y MbCas12a respectivamente.

Tabla 2A. ARNcr de Cas12a para experimentos de nucleasas

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
AACC-1	FANCF	AACC	AGTGGAGGCAAGAGGGCGGCGC	26.
AACC-2	RUNXI	AACC	AAGACAGGTCACCTGTTTCAG	27.
AACC-3	EMX1	AACC	ACACCTTCACCTGGGCCAGG	28.
AACC-4	EMX1	AACC	GGTGGCGCATTGCCACGAAG	29.

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
AATA-1	FANCF	AATA	GCATTGCAGAGAGGCGTATC	30.
AATA-2	RUNXI	AATA	TGCTGTCTGAAGCCATCGCT	31.
AATA-3	DNMT1	AATA	AGTGGCAGAGTGCTAAGGGA	32.
AATA-4	EMX1 (amplicón 2)	AATA	TGGAGCCTGCTCCAGGTGGG	33.
AATC-1	FANCF	AATC	AGTACGCAGAGAGTCGCCGT	34.
AATC-2	CFTR	AATC	CTAACTGAGACCTTACACCG	35.
AATG-1	EMX1	AATG	CGCCACCGGTTGATGTGATG	36.
ACCC-1	VEGFA	ACCC	CGGCTCTGGCTAAAGAGGGA	37.
ACCC-2	VEGFA	ACCC	CCTATTTCTGACCTCCCAAA	38.
ACCC-3	DNMT1	ACCC	AGAGGCTCAAGTGAGCAGCT	39.
ACCC-4	EMX1	ACCC	TAGTCATTGGAGGTGACATC	40.
ACCC-5	EMX1	ACCC	ACGAGGGCAGAGTGCTGCTT	41.
ACCC-6	DNMT1	ACCC	AATAAGTGGCAGAGTGCTAA	42.
AGCC-1	FANCF	AGCC	GCCCTCTTGCCTCCACTGGT	43.
AGCC-2	RUNXI	AGCC	ATCGCTTCCTCCTGAAAATG	44.
AGCC-3	RUNXI	AGCC	TCACCCCTCTAGCCCTACAT	45.
AGCC-4	EMX1 (amplicón 2)	AGCC	TGCTCCAGGTGGGGAATAAG	46.
AGTA-1	DNMT1	AGTA	ACAGACATGGACCATCAGGA	47.
AGTA-2	CFTR	AGTA	CCAGATTCTGAGCAGGGAGA	48.
AGTC-1	DNMT1	AGTC	TCCGTGAACGTTCCCTTAGC	49.
AGTC-2	CFTR	AGTC	TGTCCTGAACCTGATGACAC	50.

ES 2 992 912 T3

ATCA-1	DNMT1	ATCA	GGAAACATTAACGTAAGT	51.
ATCA-2	CFTR	ATCA	GAATCCTCTTCGATGCCATT	52.
ATCC-1	DNMT1	ATCC	TCACAGCAGCCCCCTTGAGAA	53.
ATCC-2	CFTR	ATCC	AATCAACTCTATACGAAAAT	54.
ATCC-3	DNMT1	ATCC	CCAACATGCACTGATGTTGT	55.
ATCC-4	FANCF	ATCC	ATCGGCGCTTTGGTCGGCAT	56.
ATTA-1	DNMT1	ATTA	ACGTAAGTATGTTAACAGCT	57.
ATTA-2	EMX1 (amplicón 2)	ATTA	ACATTAACAAGAAGCATTG	58.
ATTA-3	EMX1 (amplicón 2)	ATTA	TTCAAGTGGCGCAGATCTAG	59.
ATTA-4	CFTR	ATTA	GAAGGAGATGCTCCTGTCTC	60.
ATTC-1	DNMT1	ATTC	ACCGAGCAGGAGTGAGGGAA	61.
ATTC-2	EMX1 (amplicón 2)	ATTC	CCCACCTGGAGCAGGCTCCA	62.
ATTC-3	CFTR	ATTC	TGATGAGCCTTTAGAGAGAA	63.
ATTC-4	VEGFA	ATTC	CCTCTTTAGCCAGAGCCGGG	64.
ATTC-5	FANCF	ATTC	GCACGGCTCTGGAGCGGCGG	65.
ATTG-1	DNMT1	ATTG	GGTCAGCTGTTAACATCAGT	66.
ATTG-2	EMX1 (amplicón 2)	ATTG	TTATGAACCTGGGTGAAGTC	67.
ATTG-3	VEGFA	ATTG	GAATCCTGGAGTGACCCCTG	68.
ATTG-4	CFTR	ATTG	GATTGAGAATAGAATTCTTC	69.
ATTG-5	FANCF	ATTG	GAACATCCGCGAAATGATAC	70.
ATTT-1	DNMT1	ATTT	GGCTCAGCAGGCACCTGCCT	71.
ATTT-2	EMX1 (amplicón 2)	ATTT	GCTTTCCACCCACCTTTCCC	72.
ATTT-3	VEGFA	ATTT	CTGACCTCCCAAACAGCTAC	73.
ATTT-4	CFTR	ATTT	CTTCTTTCTGCACTAAATTG	74.
ATTT-5	FANCF	ATTT	CGCGGATGTTCCAATCAGTA	75.
CACC-1	FANCF	CACC	GTGCGCCGGGCTTGCAAGT	76.
CACC-2	RUNXI	CACC	GAGGCATCTCTGCACCGAGG	77.
CCCC-1	FANCF	CCCC	GCCCAAAGCCGCCCTCTTGC	78.
CCCC-2	RUNXI	CCCC	GCCTTCAGAAGAGGGTGCAT	79.

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
CCCC-3	DNMT1	CCCC	AGAGGGTTCTAGACCCAGAG	80.
CCCC-4	DNMT1	CCCC	AGGGCCAGCCAGCAGCCAA	81.
CGCA-1	FANCF	CGCA	CGGCTCTGGAGCGGCGGCTG	82.
CGCA-2	EMX1	CGCA	TTGCCACGAAGCAGGCCAAT	83.
CGCC-1	FANCF	CGCC	GCTCCAGAGCCGTGCGAATG	84.
CGCC-2	EMX1	CGCC	ACCGGTTGATGTGATGGGAG	85.
CGCC-3	FANCF	CGCC	ACATCCATCGGCGCTTTGGT	86.
CGCC-4	FANCF	CGCC	GATGGATGTGGCGCAGGTAG	87.
CGTC-1	FANCF	CGTC	AGCACCTGGGACCCCGCCAC	88.
CGTC-2	FANCF	CGTC	TCCAAGGTGAAAGCGGAAGT	89.
CTCA-1	DNMT1	CTCA	AACGGTCCCCAGAGGGTTCT	90.
CTCA-2	CFTR	CTCA	AAACTCATGGGATGTGATTC	91.
CTCC-1	DNMT1	CTCC	GTGAACGTTCCCTTAGCACT	92.
CTCC-2	CFTR	CTCC	TTCTAATGAGAAACGGTGTA	93.
CTCC-3	FANCF	CTCC	ACTGGTTGTGCAGCCGCCGC	94.
CTCC-4	FANCF	CTCC	AGAGCCGTGCGAATGGGGCC	95.

CTCT-1	DNMT1	CTCT	GGGGAACACGCCCCGGTGTCA	96.
CTTA-1	DNMT1	CTTA	TTGGGTCAGCTGTTAACATC	97.
CTTA-2	EMX1 (amplicón 2)	CTTA	TTCCCCACCTGGAGCAGGCT	98.
CTTA-3	RUNXI	CTTA	CTAATCAGATGGAAGCTCTT	99.
CTTA-4	CFTR	CTTA	CACCGTTTCTCATTAGAAGG	100.
CTTC-1	FANCF	CTTC	CGCTTTCACCTTGAGACGG	101.
CTTC-2	EMX1 (amplicón 2)	CTTC	ACCCAGGTTTATAACAATGT	102.
CTTC-3	VEGFA	CTTC	TCCCCGCTCCAACGCCCTCA	103.
CTTC-4	CFTR	CTTC	TAATGAGAAACGGTGTAAGG	104.
CTTC-5	FANCF	CTTC	GCGCACCTCATGGAATCCCT	105.
CTTG-1	DNMT1	CTTG	ACAGGCGAGTAACAGACATG	106.
CTTG-2	EMX1 (amplicón 2)	CTTG	TTAATGTTAATAACTTGCTT	107.
CTTG-3	CFTR	CTTG	GTAACTGAGTGTGTCATCA	108.
CTTG-4	RUNXI	CTTG	GGGAGTCCCAGAGGTATCCA	109.
CTTT-1	DNMT1	CTTT	GGTCAGGTTGGCTGCTGGGC	110.
CTTT-2	EMX1 (amplicón 2)	CTTT	CCCTGGCCTACCTCACTGGC	111.
CTTT-3	VEGFA	CTTT	AGCCAGAGCCGGGGTGTGCA	112.
CTTT-4	CFTR	CTTT	AGAGAGAAGGCTGTCTTAG	113.
CTTT-5	FANCF	CTTT	GGTCGGCATGGCCCCATTCTG	114.
GCCC-1	DNMT1	GCCC	GGTGTACGCCACTTGACAG	115.
GCCC-2	CFTR	GCCC	CACGCTTCAGGCACGAAGGA	116.
GTCA-1	DNMT1	GTCA	CGCCACTTGACAGGCGAGTA	117.
GTCA-2	CFTR	GTCA	TCAGGTTTCAGGACAGACTGC	118.
GTCC-1	DNMT1	GTCC	CCAGAGGGTTCTAGACCCAG	119.
GTCC-2	CFTR	GTCC	AGGAGACAGGAGCATCTCCT	120.
GTCC-3	FANCF	GTCC	CAGGTGCTGACGTAGGTAGT	121.
GTCC-4	EMX1	GTCC	TCCCCATTGGCCTGCTTCGT	122.
GTTA-1	DNMT1	GTTA	CTCGCCTGTCAAGTGCGGTG	123.
GTTA-2	EMX1 (amplicón 2)	GTTA	TGAACCTGGGTGAAGTCCCA	124.
GTTA-3	EMX1 (amplicón 2)	GTTA	ATGTTAATAACTTGCTTCAA	125.
GTTA-4	CFTR	GTTA	ACCAAGGTCAGAACATTAC	126.
GTTT-1	DNMT1	GTTT	CCTTAGCACTCTGCCACTTA	127.
GTTT-2	EMX1 (amplicón 2)	GTTT	ATTTGTCCAGAGGAAACCAC	128.
GTTT-3	RUNXI	GTTT	CCTGTCTTGTGTGTGAGAGG	129.

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
GTTT-4	CFTR	GTTT	AGGACAGACTGCCTCCTTCG	130.
GTTG-1	DNMT1	GTTG	GGGATTCCTGGTGCCAGAAA	131.
GTTG-2	EMX1 (amplicón 2)	GTTG	GGACTTCACCCAGGTTTATA	132.
GTTG-3	VEGFA	GTTG	AGGGCGTTGGAGCGGGGAGA	133.
GTTG-4	CFTR	GTTG	ATTGGATTGAGAATAGAATT	134.
GTTG-5	FANCF	GTTG	TGCAGCCGCCGCTCCAGAGC	135.
GTTT-1	DNMT1	GTTT	CCTGATGGTCCATGTCTGTT	136.
GTTT-2	EMX1 (amplicón 2)	GTTT	GACTTGGGATAGTGGAATAG	137.
GTTT-3	VEGFA	GTTT	GGGAGGTCAGAAATAGGGGG	138.
GTTT-4	CFTR	GTTT	CTCATTAGAAGGAGATGCTC	139.
GTTT-5	RUNXI	GTTT	CACCTCGGTGCAGAGATGCC	140.

ES 2 992 912 T3

TACA-1	RUNXI	TACA	TCTCTCTTTCTTCTCCCCTC	141.
TACA-2	RUNXI	TACA	GGCAAAGCTGAGCAAAAGTA	142.
TACA-3	EMX1	TACA	AACGGCAGAAGCTGGAGGAG	143.
TACA-4	RUNXI	TACA	AGACCAGCATGTACTCACCT	144.
TACC-1	DNMT1	TACC	CACGTTTCGTGGCCCCATCTT	145.
TACC-2	CFTR	TACC	AGATTCTGAGCAGGGAGAGG	146.
TACC-3	EMX1 (amplicón 2)	TACC	TCACTGGCCCCACCCAGAG	147.
TACC-4	FANCF	TACC	TGCGCCACATCCATCGGCGC	148.
TATA-1	CFTR	TATA	GAGTTGATTGGATTGAGAAT	149.
TATA-2	CFTR	TATA	TTCAAGAAGGTTATCTCAAG	150.
TATA-3	CFTR	TATA	TATTCAAGAAGGTTATCTCA	151.
TATA-4	VEGFA (amplicón 2)	TATA	GACATGTCCCATTGTGGGA	152.
TATC-1	CFTR	TATC	GCCTCTCCCTGCTCAGAATC	153.
TATC-2	CFTR	TATC	TCAAGAAACTGGCTTGGA	154.
TATC-3	CFTR	TATC	CAGTTCAGTCAAGTTTGCCT	155.
TATC-4	EMX1 (amplicón 2)	TATC	CCAAGTCAAACCTTCTCTTCA	156.
TATG-1	VEGFA (amplicón 2)	TATG	TTCGGGTGCTGTGAACCTCC	157.
TATG-2	EMX1 (amplicón 2)	TATG	AACCTGGGTGAAGTCCCAAC	158.
TATG-3	VEGFA	TATG	TAGCTGTTTGGGAGGTCAGA	159.
TATG-4	CFTR	TATG	GGACATTTTCAGAACTCCAA	160.
TATT-1	DNMT1	TATT	GGGTGAGCTGTTAACATCAG	161.
TATT-2	VEGFA	TATT	TCTGACCTCCCAAACAGCTA	162.
TATT-3	CFTR	TATT	CTCAATCCAATCAACTCTAT	163.
TATT-4	FANCF	TATT	GGTCGAAATGCATGTCAATC	164.
TCCA-1	DNMT1	TCCA	TGTCTGTTACTCGCCTGTCA	165.
TCCA-2	CFTR	TCCA	GGAGACAGGAGCATCTCCTT	166.
TCCA-3	VEGFA	TCCA	GTCCCAAATATGTAGCTGTT	167.
TCCC-1	DNMT1	TCCC	CAGAGGGTTCTAGACCCAGA	168.
TCCC-2	CFTR	TCCC	CAAACCTCTCCAGTCTGTTTA	169.
TCCC-3	DNMT1	TCCC	GTCACCCCTGTTTCTGGCAC	170.
TCCC-4	FANCF	TCCC	AGGTGCTGACGTAGGTAGTG	171.
TCCC-5	VEGFA	TCCC	TCTTTAGCCAGAGCCGGGGT	172.
TCCG-1	DNMT1	TCCG	TGAACGTTCCCTTAGCACTC	173.
TCCG-2	FANCF	TCCG	AGCTTCTGGCGGTCTCAAGC	174.
TCCG-3	VEGFA	TCCG	CACGTAACCTCACTTTCCCTG	175.
TCCT-1	DNMT1	TCCT	GATGGTCCATGTCTGTTACT	176.
TGCA-1	DNMT1	TGCA	CACAGCAGGCCTTTGGTCAG	177.
TGCA-2	CFTR	TGCA	GAAAGAAGAAATTCAATCCT	178.
TGCC-1	DNMT1	TGCC	ACTTATTGGGTCAGCTGTTA	179.

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
TGCC-2	CFTR	TGCC	TCGCATCAGCGTGATCAGCA	180.
TGCC-3	FANCF	TGCC	TCCACTGGTTGTGCAGCCGC	181.
TGCC-4	FANCF	TGCC	GACCAAAGCGCCGATGGATG	182.
TGTA-1	RUNXI	TGTA	ATGAAATGGCAGCTTGTTTC	183.
TGTA-2	EMX1	TGTA	CTTTGTCCTCCGGTTCTGGA	184.

ES 2 992 912 T3

TGTA-3	Sitio coincidente 5	TGTA	CCTCACCCTGACATTAATT	185.
TGTA-4	Sitio coincidente 5	TGTA	ACCACAGTCAAGTAGTTAAT	186.
TGTA-5	CFTR	TGTA	AGGTCTCAGTTAGGATTGAA	187.
TGTC-1	FANCF	TGTC	AATCTCCCAGCGTCTTTATC	188.
TGTC-2	RUNXI	TGTC	TTGTTTGTGAGAGGAATTCA	189.
TGTC-3	EMX1 (amplicón 2)	TGTC	CAGAGGAAACCACTGTTGGG	190.
TGTC-4	EMX1 (amplicón 2)	TGTC	TATTCCACTATCCCAAGTCA	191.
TGTC-5	EMX1 (amplicón 2)	TGTC	CCTCACCCTCTCCCTGTGA	192.
TGTG-1	FANCF	TGTG	CAGCCGCCGCTCCAGAGCCG	193.
TGTG-2	RUNXI	TGTG	AGAGGAATTCAAAGTGAAGC	194.
TGTG-3	FANCF	TGTG	GCGCAGGTAGCGCGCCCACT	195.
TGTG-4	EMX1	TGTG	GTTCCAGAACCAGGAGGACAA	196.
TGTG-5	EMX1	TGTG	ATGGGAGCCCTTCTTCTTCT	197.
TTAC-2	RUNXI	TTAC	AGGCAAAGCTGAGCAAAAGT	198.
TTAC-3	EMX1 (amplicón 2)	TTAC	TCATCTCTGCCAGACCACCT	199.
TTAC-4	Sitio coincidente 5	TTAC	TGATTCTGGGGTCAACATCT	200.
TTAC-5	CFTR	TTAC	AAATGAATGGCATCGAAGAG	201.
TTAC-6	CFTR	TTAC	ACCGTTTCTCATTAGAAGGA	202.
TTAC-7	Sitio coincidente 5	TTAC	TAGGGCAATAAGCAACACCT	203.
TTCA-1	DNMT1	TTCA	GTCTCCGTGAACGTTCCCTT	204.
TTCA-2	EMX1 (amplicón 2)	TTCA	CCCAGGTTTATAACAATGTT	205.
TTCA-3	VEGFA	TTCA	CCCAGCTTCCCTGTGGTGGC	206.
TTCA-4	CFTR	TTCA	ATCCTAACTGAGACCTTACA	207.
TTCA-5	FANCF	TTCA	CCTTGGAGACGGCGACTCTC	208.
TTCC-1	DNMT1	TTCC	TGATGGTCCATGTCTGTTAC	209.
TTCC-2	EMX1 (amplicón 2)	TTCC	CTGGCCTACCTCACTGGCCC	210.
TTCC-3	VEGFA	TTCC	AAAGCCCATTCCCTCTTTAG	211.
TTCC-4	CFTR	TTCC	ATTGTGCAAAAGACTCCCTT	212.
TTCC-5	FANCF	TTCC	GAGCTTCTGGCGGTCTCAAG	213.
TTCCG-1	DNMT1	TTCCG	TGGCCCCATCTTTCTCAAGG	214.
TTCCG-2	VEGFA	TTCCG	AGAGTGAGGACGTGTGTGTC	215.
TTCCG-3	RUNXI	TTCCG	GAGCGAAAACCAAGACAGGT	216.
TTCCG-4	CFTR	TTCCG	ACCAATTTAGTGCAGAAAGA	217.
TTCCG-5	FANCF	TTCCG	CACGGCTCTGGAGCGGCGGC	218.
TTCT-1	DNMT1	TTCT	GCCCTCCCGTCACCCCTGTT	219.
TTCT-2	EMX1 (amplicón 2)	TTCT	GCCCTTTACTCATCTCTGCC	220.
TTCT-3	VEGFA	TTCT	GACCTCCCAAACAGCTACAT	221.
TTCT-4	CFTR	TTCT	TTCGACCAATTTAGTGCAGA	222.
TTCT-5	FANCF	TTCT	GGCGGTCTCAAGCACTACCT	223.
TTTA-1	DNMT1	TTTA	TTTCCCTTCAGCTAAAATAA	224.
TTTA-2	DNMT1	TTTA	TTTGTAGCTGAAGGGAAATAA	225.
TTTA-3	FANCF	TTTA	TCCGTGTTCCCTGACTCTGG	226.
TTTA-4	RUNXI	TTTA	CCTTCGGAGCGAAAACCAAG	227.
TTTA-5	Sitio coincidente 5	TTTA	GGATGCCACTAAAAGGGAAA	228.
TTTA-6	Sitio coincidente 1	TTTA	GATTGAAGGAAAAGTTACAA	229.

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
TTTC-1	DNMT1	TTTC	CCTCACTCCTGCTCGGTGAA	230.
TTTC-2	DNMT1	TTTC	CTGATGGTCCATGTCTGTTA	231.
TTTC-3	EMX1	TTTC	TCATCTGTGCCCCCTCCCTCC	232.
TTTC-4	FANCF	TTTC	ACCTTGGAGACGGCGACTCT	233.
TTTC-5	RUNXI	TTTC	GCTCCGAAGGTAAAAGAAAT	234.
TTTC-6	RUNXI	TTTC	AGCCTCACCCCTCTAGCCCT	235.
TTTC-7	RUNXI	TTTC	TTCTCCCCTCTGCTGGATAC	236.
TTTC-8	FANCF	TTTC	CGAGCTTCTGGCGGTCTCAA	237.
TTTG-1	DNMT1	TTTG	AGGAGTGTTCACTCTCCGTG	238.
TTTG-2	DNMT1	TTTG	GCTCAGCAGGCACCTGCCTC	239.
TTTG-3	EMX1	TTTG	TCCTCCGGTTCTGGAACCAC	240.
TTTG-4	EMX1	TTTG	TGGTTGCCACCCCTAGTCAT	241.
TTTG-5	EMX1	TTTG	TACTTTGTCTCCGGTTCTG	242.
TTTG-6	FANCF	TTTG	GGCGGGGTCCAGTTCCGGGA	243.
TTTG-7	FANCF	TTTG	GTCGGCATGGCCCCATTGCG	244.
TTTT-1	DNMT1	TTTT	ATTTCCCTTCAGCTAAAATA	245.
TTTT-2	RUNXI	TTTT	CAGGAGGAAGCGATGGCTTC	246.
TTTT-3	FANCF	TTTT	CCGAGCTTCTGGCGGTCTCA	247.
TTTT-4	CFTR	TTTT	CGTATAGAGTTGATTGGATT	248.
TTTT-5	CFTR	TTTT	GAGCTAAAGTCTGGCTGTAG	249.

Tabla 2B. ARNcr de Cas12a para experimentos de activación de genes

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:	Distancia de TSS (bp) †	Cadena
AR-TTIV- a-1	AR	TTTG	AGAGTCTGGATGAGAAATG C	250.	639	C
AR-TTIV- a-2	AR	TTTC	TACCCTCTTCTCTGCCTTT C	251.	588	T
AR-TTIV- a-3	AR	TTTG	CTCTAGGAACCCCTCAGCCC C	252.	550	T
AR-TTIV- b-1	AR	TTTC	TCCAAAGCCACTAGGCAGG C	253.	141	C
AR-TTIV- b-2	AR	TTTA	GGAAAGCAGGAGCTATTCA G	254.	231	C
AR-TTIV- b-3	AR	TTTG	GAACCAAATTTGGTGAGTG C	255.	296	C
AR-ATTIV-1	AR	ATTC	AGGAAGCAGGGTCTCTCCA G	256.	142	C
AR-ATTIV-2	AR	ATTG	GGCTTTGGAACCAAATTTG G	257.	303	C
AR-ATTIV-3	AR	ATTC	CGTCATAGGGATAGATCGG G	258.	508	T
AR-CTIV-1	AR	CTTG	TTTCTCCAAAGCCACTAGG C	259.	145	C
AR-CTIV-2	AR	CTTC	CTGAATAGCTCCTGCTTTC C	260.	227	T
AR-CTIV-3	AR	CTTA	TCAGTCCTGAAAAGAACCC C	261.	398	C

5

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO :	Distancia de TSS (bp) †	Cadena
AR-GTIV-1	AR	GTTG	CATTTGCTCTCCACCTCCC A	262.	9	C
AR-GTIV-2	AR	GTTA	GCGCGCGGTGAGGGGAGGG G	263.	117	C
AR-GTIV-3	AR	GTTT	CAAAGCCCAATCTAAAAAA C	264.	312	T

ES 2 992 912 T3

AR-TTCV-1	AR	TTCA	GGAAGCAGGGGTCCTCCAG G	265.	212	C
AR-TTCV-2	AR	TTCC	TGGAGGCCAGCACTCACCAA	266.	283	T
AR-TTCV-3	AR	TTCA	GGACTGATAAGAGCGCGCA G	267.	407	T
AR-CTCC-1	AR	CTCC	AAAGCCACTAGGCAGGCGT T	268.	138	C
AR-CTCC-2	AR	CTCC	AGGAAATCTGGAGCCCTGG C	269.	268	C
AR-CTCC-3	AR	CTCC	CTCCCTCGCCTCCACCCTG T	270.	338	C
AR-TCCC-1	AR	TCCC	GCCCCACCGGGCCGCGCCT C	271.	48	T
AR-TCCC-2	AR	TCCC	CTCACC GCGCGCTAACGCCT	272.	121	T
AR-TCCC-3	AR	TCCC	TCGCCTCCACCCTGTTGGT T	273.	333	C
HBB-TTTV-1	HBB	TTTG	TACTGATGGTATGGGGCCA A	274.	203	C
HBB-TTTV-2	HBB	TTTG	AAGTCCAACCTCCTAAGCCA G	275.	150	C
HBB-TTTV-3	HBB	TTTG	CAAGTGATTTACGTAATA T	276.	i 248	T
HBB-ATTV-1	HBB	ATTG	GCCAACCCTAGGGTGTGGC T	277.	71	T
HBB-ATTV-2	HBB	ATTG	CTACTAAAAACATCCTCCT T	278.	226	T
HBB-ATTV-3	HBB	ATTG	GGAAAACGATCTTCAATAT G	279.	293	T
HBB-CTTV-1	HBB	CTTA	GACCTCACCTGTGGAGCC A	280.	90	C
HBB-CTTV-2	HBB	CTTA	GGAGTTGGACTTCAAACCC T	281.	154	T
HBB-CTTV-3	HBB	CTTA	CCAAGCTGTGATTCCAAAT A	282.	269	C
HBB-TATV-1	HBB	TATG	CCCAGCCCTGGCTCCTGCC C	283.	28	T
HBB-TATV-2	HBB	TATC	TCTTGCCCCATACCATCA G	284.	197	T
HBB-TATV-3	HBB	TATC	CCAAAGCTGAATTATGGTA G	285.	369	C
HBB-TGTV-1	HBB	TGTC	ATCACTTAGACCTCACCT G	286.	98	C

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:	Distancia de TSS (bp) ‡	Cadena
HBB-TGTV-2	HBB	TGTA	CTGATGGTATGGGGCCAAG A	287.	203	C
HBB-TGTV-4	HBB	TGTA	GATGGATCTCTCCTGCGT C	288.	393	T
HBB-TTCV-1	HBB	TTCA	AACCCTCAGCCCTCCCTCT A	289.	167	T
HBB-TTCV-2	HBB	TTCC	AAATATTACGTAATAACAC T	290.	254	C
HBB-TTCV-3	HBB	TTCA	GCTTTGGGATATGTAGATG G	291.	378	T
HBB-CTCC-1	HBB	CTCC	CTGCTCCTGGGAGTAGATT G	292.	51	T
HBB-CTCC-2	HBB	CTCC	CTCTAAGATATATCTTTG G	293.	183	T
HBB-CTCC-3	HBB	CTCC	AGAATATGCAAAATACTTA C	294.	417	T
HBB-TACC-1	HBB	TACC	TGTCCTTGGCTCTTCTGGC A	295.	126	T
HBB-TACC-2	HBB	TACC	ATCAGTACAAATTGCTACT A	296.	212	T
HBB-TACC-3	HBB	TACC	ATAATTCAGCTTTGGGATA T	297.	370	T
NPY1R-TTTV-1	NPY1 R	TTTC	AAGCCTCGGGAAACTGCCC T	298.	256	C
NPY1R-TTTV-2	NPY1 R	TTTC	TTTGTTTGCAGGTCAAGTGC C	299.	299	T
NPY1R-TTTV-3	NPY1 R	TTTG	GGCTGGCGCTCGAGCTCTC C	300.	350	C
NPY1R-ATTV-1	NPY1 R	ATTC	CTGGTTTGGGCTGGCGCTC G	301.	382	C
NPY1R-ATTV-2	NPY1 R	ATTA	GTGCCATTATTGTGGCGAA T	302.	407	C
NPY1R-ATTV-3	NPY1 R	ATTC	TCGGCACTGGCGTGAGAGT T	303.	1 464	C
NPY1R-CTTV-1	NPY1 R	CTTC	CCCGGAGTCGAGGACTGTG G	304.	230	C
NPY1R-CTTV-2	NPY1 R	CTTC	GGCCACAAGATGGCACTGA C	305.	314	C
NPY1R-CTTV-3	NPY1 R	CTTA	TAAAGTGAGGAAAACAAAT T	306.	485	C
NPY1R-TTCV-1	NPY1 R	TTCC	CCGGAGTCGAGGACTGTGG G	307.	229	C
NPY1R-TTCV-2	NPY1 R	TTCC	GCCACAAGATGGCACTGAC C	308.	313	C
NPY1R-TTCV-3	NPY1 R	TTCC	CAGCGAGCCCTTTGATTCC T	309.	376	C
NPY1R-CTCC-1	NPY1 R	CTCC	GGGAAGGCAGGGCAGTTT C	310.	243	T

ES 2 992 912 T3

NPY1R-CTCC-2 NPY1 R CTCC AGCCGGGTATGACTTCGGC C 311. 330 C

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:	Distancia de TSS (bp) †	Cadena
NPY1R-CTCC-3	NPY1 R	CTCC	TTTCTTTGGCCCACTGAGA A	312.	554	T
VEGFA-TTTV-1	VEGF A	TTTC	AGGCTGTGAACCTTGGTGG G	313.	200	C
VEGFA-TTTV-2	VEGF A	TTTC	CTGCTCCCTCCTCGCCAAT G	314.	274	C
VEGFA-TTTV-3	VEGF A	TTTG	CTAGGAATATTGAAGGGGG C	315.	338	T
VEGFA-ATTV-1	VEGF A	ATTG	CGGCGGGCTGCGGGCCAGG C	316.	159	C
VEGFA-ATTV-2	VEGF A	ATTA	CCCATCCGCCCCCGGAAAC T	317.	274	T
VEGFA-ATTV-3	VEGF A	ATTC	CTAGCAAAGAGGGAACGGC T	318.	326	C
VEGFA-CTTV-1	VEGF A	CTTC	CCCTTCATTGCGGCGGGCT G	319.	114	T
VEGFA-CTTV-2	VEGF A	CTTC	CCCTTCATTGCGGCGGGCT G	320.	169	C
VEGFA-CTTV-3	VEGF A	CTTC	CCCTGCCCCCTTCAATATT C	321.	346	C
VEGFA-GTTV-1	VEGF A	GTTC	ACAGCCTGAAAATTACCCA T	322.	209	T
VEGFA-GTTV-2	VEGF A	GTTA	CGTGCGGACAGGGCCTGAG A	323.	303	T
VEGFA-GTTV-3	VEGF A	GTTG	GAGCGGGGAGAAGGCCAGG G	324.	435	C
VEGFA-TTCV-1	VEGF A	TTCC	ACACGCGGCTCGGGCCCGG G	325.	115	T
VEGFA-TTCV-2	VEGF A	TTCA	GGCTGTGAACCTTGGTGGG G	326.	199	C
VEGFA-TTCV-3	VEGF A	TTCC	TGCTCCCTCCTCGCCAATG C	327.	213	C
VEGFA-TTCV-4	VEGF A	TTCC	CCTTCATTGCGGCGGGCTG C	328.	185	C
VEGFA-TTCV-5	VEGF A	TTCC	CCTGCCCCCTTCAATATTC C	329.	362	C
VEGFA-TCCC-1	VEGF A	TCCC	CTTCATTGCGGCGGGCTGC G	330.	i 167	C
VEGFA-TCCC-2	VEGF A	TCCC	TCCTCGCCAATGCCCCGCG G	331.	266	C
VEGFA-TCCC-3	VEGF A	TCCC	CTGCCCCCTTCAATATTCC T	332.	344	C
VEGFA-	VEGF		TCGCCAATGCCCCGCGGGC	333.		

CTCC-1	A	CTCC	G	263	C
VEGFA-CTCC-2	VEGF	CTCC	CTCCTCGCCAATGCCCCGC	334.	
	A	CTCC	G	267	C

(continúa)

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:	Distancia de TSS (bp) ‡	Cadena
VEGFA-CTCC-3	VEGF	CTCC	TTTCTTTGGCCCACTGAGA	335.		
	A	CTCC	G		407	C

C, codificación; T, plantilla; ‡ medido desde el TSS hasta la posición -4 del PAM para guías de plantilla-cadena o el 20° nt del espaciador para guías de codificación-cadena

Tabla 2C: ARNsg de SpCas9 para experimentos de activación de genes VEGFA

ID guía	3 nt PAM	Espaciador 20nt	SEQ ID NO:	Distancia de TSS (bp)	Cadena
VEGFA-NGG-a-1	AGG	GTGTGCAGACGGCAGTCA C T	336.	571	codificación
VEGFA-NGG-a-2	AGG	GAGCAGCGTCTTCGAGAG T G	337.	509	codificación
VEGFA-NGG-a-3	TGG	GGTGAGTGAGTGTGTGCG T G	338.	469	codificación
VEGFA-NGG-b-4	AGG	GGGGCGGATGGGTAATTTT C	339.	217	codificación
VEGFA-NGG-b-5	AGG	GGCATTGGCGAGGAGGGA G C	340.	272	Plantilla
VEGFA-NGG-b-6	AGG1	GCAAAGAGGGAACGGCTC T C	341.	320	codificación

‡, medido desde TSS hasta la posición -3 de PAM para guías de codificación-cadena o el 20° nt del espaciador para guías de plantilla-cadena

5

Tabla 2D: ARNcr de Cas12a para experimentos de editor de bases

ID de ARNcr	nombre del gen	4 nt PAM	Espaciador 20nt	SEQ ID NO:
TTTA-3	FANCF	TTTA	TCCGTGTTCTTGACTCTGG	342.
TTTC-1	DNMT1	TTTC	CCTCACTCCTGCTCGGTGAA	343.
TTTC-3	EMX1	TTTC	TCATCTGTGCCCCCTCCCTCC	344.
TTTC-6	RUNX1	TTTC	AGCCTCACCCCTCTAGCCCT	345.
TTTC-7	RUNX1	TTTC	TTCTCCCCTCTGCTGGATAC	346.
TTTC-8	FANCF	TTTC	CGAGCTTCTGGCGGTCTCAA	347.
TTTG-4	EMX1	TTTG	TGGTTGCCACCCTAGTCAT	348.
TTTG-7	FANCF	TTTG	GTCGGCATGGCCCCATTCGC	349.

Tabla 3A. Oligonucleótidos utilizados en este estudio - Para experimentos de T7E1 y RFLP

descripción	secuencia	SEQ ID NO:
cebador de PCR directo para amplificar locus DNMT1 en células humanas	CCAGAATGCACAAAGTACTGCAC	350.
cebador de PCR inverso para amplificar locus DNMT1 en células humanas	GCCAAAGCCCCGAGAGAGTGCC	351.

cebador de PCR directo para amplificar locus CFTR en células humanas	GCTGTGTCTGTAACTGATGGCTAACA	352.
cebador de PCR inverso para amplificar locus CFTR en células humanas	TTGCATTCTACTCAATTGCATTCTGTGGG	353.
cebador de PCR directo para amplificar locus EMX1 en células humanas	GGAGCAGCTGGTCAGAGGGG	354.

(continúa)

descripción	secuencia	SEQ ID NO:
cebador de PCR inverso para amplificar locus EMX1 en células humanas	CCATAGGGAAGGGGGACACTGG	355.
cebador de PCR directo para amplificar locus EMX1 (amplicón 2) en células humanas	CTGCCTCCTATTACACACTTACGGG	356.
cebador de PCR inverso para amplificar locus EMX1 (amplicón 2) en células humanas	CTCTGTTGGTGGAACTCCCTGACC	357.
cebador de PCR directo para amplificar locus FANCF en células humanas	GGGCCGGGAAGAGTTGCTG	358.
cebador de PCR inverso para amplificar locus FANCF en células humanas	GCCCTACATCTGCTCTCCCTCC	359.
cebador de PCR directo para amplificar locus RUNX1 en células humanas	CCAGCACAACCTACTCGCACTTGAC	360.
cebador de PCR inverso para amplificar locus RUNX1 en células humanas	CATCACCAACCCACAGCCAAGG	361.
cebador de PCR directo para amplificar locus VEGFA en células humanas	CAGCTCCACAACTTGGTGCCAAATTC	362.
cebador de PCR inverso para amplificar locus VEGFA en células humanas	CCGCAATGAAGGGGAAGCTCGAC	363.
cebador de PCR directo para amplificar locus VEGFA (amplicón 2) en células humanas	CGCTGTTCAAGTCTCTGCTAGAAGTAGG	364.
cebador de PCR inverso para amplificar locus VEGFA (amplicón 2) en células humanas	CCAGACCAGAGACCACTGGGAAG	365.
cebador de PCR directo para amplificar locus de Sitio coincidente 1 en células humanas	GACAAATGTATCATGCTATTATAAGATGTTGAC	366.
cebador de PCR inverso para amplificar locus de Sitio coincidente 1 en células humanas	CCATTTACTGAGAGTAATTATAATTGTGC	367.
cebador de PCR directo para amplificar locus de Sitio coincidente 5 en células humanas	CCAAGGACAGGAATATCTTATACCCTCTGT	368.
cebador de PCR inverso para amplificar locus de Sitio coincidente 5 en células humanas	TGTCATTGTCCTTGTCTTTAGCTACCG	369.

5 Tabla 3B. Oligonucleótidos utilizados en este estudio - Para el ensayo de determinación de PAM (PAMDA) y otros experimentos de escisión *in vitro*

descripción	Secuencia	SEQ ID NO:
cebador de PCR inverso para amplificar locus de PAM aleatorizado	CAAAACAGCCAAGCTTGCATGC	370.
cebador de PCR directo para amplificar locus de PAM aleatorizado, añadiendo el código de barras CCAT	AGCTGCCATCGGTATTTACACCCGCATACG TAC	371.
cebador de PCR directo para amplificar locus de PAM aleatorizado, añadiendo código de barras	AGCTGGCAACGGTATTTACACCCGCATACG TAC	372.
cebador de PCR directo para amplificar locus de PAM aleatorizado, añadiendo el código de barras ATGC	GCAAGCTGATGCCGGTATTTACACCCGCATACG TAC	373.
cebador de PCR directo para amplificar locus de PAM aleatorizado, añadiendo el código de barras GATG	AGCTGGATGCGGTATTTACACCCGCATACG TAC	374.
cebador de PCR directo para amplificar locus de PAM aleatorizado, añadiendo el código de barras CGAT	AGCTGCCATCGGTATTTACACCCGCATACG TAC	375.

(continúa)

descripción	secuencia	SEQ ID NO:
oligo de cadena superior para NNNNNNNN PAM agotamiento espaciador de biblioteca 1 a clonar en EcoRI/SphI de p11-lacY-wtx1	AGACCGGAATTCNNNGTNNNNNNNNNNNGGA ATCCCTTCTGCAGCACCTGGGCGCAGGTCA CGAGGCATG	376.
oligo de cadena superior para NNNNNNNN PAM agotamiento espaciador de biblioteca 2 a clonar en EcoRI/SphI de p11-lacY-wtx1	AGACCGGAATTCNNNGTNNNNNNNNNNNCTG ATGGTCCATGTCTGTTACTCGCGCAGGTCA CGAGGCATG	377.
cebador inverso para llenar en oligos de biblioteca	/5Phos/CCTCGTGACCTGCGC	378.
cadena superior para espaciador 1 con diana PAM TTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCTTTAGGAATCCCTTCTGCAGCACCT GGGCATG	379.
cadena inferior para espaciador 1 con diana PAM TTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	CCCAGGTGCTGCAGAAGGGATTCTAAAG	380.
cadena superior para espaciador 1 con diana PAM CTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCCTTAGGAATCCCTTCTGCAGCACCT GGGCATG	381.
cadena inferior para espaciador 1 con diana PAM CTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	CCCAGGTGCTGCAGAAGGGATTCTAAGG	382.
cadena superior para espaciador 1 con diana PAM ACCT a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCACCTGGAATCCCTTCTGCAGCACCT GGGCATG	383.
cadena inferior para espaciador 1 con diana PAM ACCT a clonar en EcoRI/SphI de p11-lacY-wtx1	CCCAGGTGCTGCAGAAGGGATTCCAGGTG	384.
cadena superior para espaciador 2 con diana PAM TTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCTTTACTGATGGTCCATGTCTGTTAC TCGCATG	385.

cadena inferior para espaciador 2 con diana PAM TTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	CGAGTAACAGACATGGACCATCAGTAAAG	386.
cadena superior para espaciador 2 con diana PAM CTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCCTTACTGATGGTCCATGTCTGTTAC TCGCATG	387.
cadena inferior para espaciador 2 con diana PAM CTTA a clonar en EcoRI/SphI de p11-lacY-wtx1	CGAGTAACAGACATGGACCATCAGTAAGG	388.
cadena superior para espaciador 2 con diana PAM ACCT a clonar en EcoRI/SphI de p11-lacY-wtx1	AATTCACCTCTGATGGTCCATGTCTGTTAC TCGCATG	389.
cadena inferior para espaciador 2 con diana PAM ACCT a clonar en EcoRI/SphI de p11-lacY-wtx1	CGAGTAACAGACATGGACCATCAGAGGTG	390.

Tabla 3C. Oligonucleótidos utilizados en este estudio - Para experimentos RT-qPCR de activador

descripción	secuencia	SEQ ID NO:
cebador RT-qPCR directo para el gen NPY1R humano	ATGGTGAGCAGAGTGCCCTATC	391
cebador RT-qPCR inverso para el gen NPY1R humano	ATGGTCCCTGGCAGTCTCCAAA	392
cebador RT-qPCR directo para el gen AR humano	CCATCGGACTCTCATAGTTGTC	393
cebador RT-qPCR inverso para el gen AR humano	GACCTGTACTTATTGTCTCTCATC	394
cebador RT-qPCR directo para el gen HBB humano	GCACGTGGATCCTGAGAACT	395
cebador RT-qPCR inverso para el gen HBB humano	ATTGGACAGCAAGAAAGCGAG	396
cebador RT-qPCR directo para el gen HPRT1 humano	CATTATGCTGAGGATTTGGAAAGG	397

5

(continúa)

cebador RT-qPCR inverso para el gen HPRT1 humano	CTTGAGCACACAGAGGGCTACA	398.
--	------------------------	------

Tabla 3D. Oligonucleótidos utilizados en este estudio - Para experimentos de secuenciación profunda de editores de bases

descripción	secuencia	SEQ ID NO:
cebador PCR directo para amplificar el sitio 3 de PAM TTTA en células humanas	GCTCCAGAGCCGTGCGAATGG	399.
cebador PCR inverso para amplificar el sitio 3 de PAM TTTA en células humanas	GCCCTACATCTGCTCTCCCTCC	400.
cebador PCR directo para amplificar el sitio 1 de PAM TTTC en células humanas	CAGCTGACCCAATAAGTGGCAGAGTG	401.
cebador PCR inverso para amplificar el sitio 1 de PAM TTTC en células humanas	TCAGTTGGCTGCTGGGCTGG	402.
cebador PCR directo para amplificar el sitio 3 de PAM TTTC en células humanas	CCCCAGTGGCTGCTCTGGG	403.
cebador PCR inverso para amplificar el sitio 3 de PAM TTTC en células humanas	CATCGATGTCCTCCCCATTGGC	404.
cebador PCR directo para amplificar el sitio 6 de PAM TTTC en células humanas	GCTGTCTGAAGCCATCGCTTCC	405.

cebador PCR inverso para amplificar el sitio 6 de PAM TTTC en células humanas	CAGAGGTATCCAGCAGAGGGGAG	406.
cebador PCR directo para amplificar el sitio 7 de PAM TTTC en células humanas	CCTTCGGAGCGAAAACCAAGACAG	407.
cebador PCR inverso para amplificar el sitio 7 de PAM TTTC en células humanas	CAGGCAGGACGAATCACACTGAATG	408.
cebador PCR directo para amplificar el sitio 8 de PAM TTTC en células humanas	GCTCCAGAGCCGTGCGAATGG	409.
cebador PCR inverso para amplificar el sitio 8 de PAM TTTC en células humanas	GCACCTCATGGAATCCCTTCTGC	410.
cebador PCR directo para amplificar el sitio 4 de PAM TTTC en células humanas	GAAGCTGGAGGAGGAAGGGC	411.
cebador PCR inverso para amplificar el sitio 4 de PAM TTTC en células humanas	CAGCAGCAAGCAGCACTCTGC	412.
cebador PCR directo para amplificar el sitio 7 de PAM TTTC en células humanas	GCCCTCTTGCCTCCACTGGTTG	413.
cebador PCR inverso para amplificar el sitio 7 de PAM TTTC en células humanas	CCAATAGCATTGCAGAGAGGCGT	414.

Condiciones de cultivo celular.

- Se cultivaron células humanas U2OS (de Toni Cathomen, Friburgo) y HEK293 (Invitrogen) en Modified Eagle Medium de Advanced Dulbecco (A-DMEM) y DMEM, respectivamente, suplementados con un 10% de FBS inactivado por calor, y 1% de penicilina y estreptomycin, y 2 mM de GlutaMax (con la excepción de que las células HEK293 cultivadas para experimentos analizados mediante RT-qPCR utilizan medios que contienen un 0,1% de penicilina y estreptomycin que carecen de GlutaMax). Todos los reactivos de cultivo celular se adquirieron a Life Technologies, y las células se cultivaron a 37 °C en un 5% de CO₂. El sobrenadante del medio se analizó cada dos semanas para detectar la presencia de micoplasma, y las identidades de las líneas celulares se confirmaron mediante perfiles STR (ATCC). Salvo que se indique lo contrario, las transfecciones de control negativo incluyeron la expresión de Cas12a y plásmidos U6-null.

Evaluación de la edición de genes y bases mediante T7E1 o secuenciación profunda.

- Para los experimentos con nucleasas y editores de bases, los plásmidos de expresión Cas12a y ARNcr (500 ng y 250 ng, respectivamente) se electroporaron en aproximadamente 2×10^5 células U2OS mediante el programa DN-100 con el SE Cell Line Nucleofector Kit utilizando un 4D-Nucleofector (Lonza). El ADN genómico (ADNg) se extrajo aproximadamente 72 o 120 horas después de la nucleofección (para experimentos con nucleasas o edición de bases, respectivamente) utilizando el kit de aislamiento de ácido nucleico Agencourt DNAdvance (Beckman Coulter), o por lisis personalizada y extracción con perlas paramagnéticas. Perlas paramagnéticas preparadas de forma similar a la descrita previamente (Rohland et al., Genome Res., 2012, 22:939-46) (GE Healthcare Sera-Mag SpeedBeads (Fisher Scientific) lavadas en 0,1 × de TE y suspendidas en 20% de PEG-8000 (p/v), 1,5 M NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8, y 0,05% Tween20). Para la lisis celular, se eliminó el sobrenadante del medio, se realizó un lavado con 500 µL de PBS y las células se trataron con 200 µL de tampón de lisis (100 mM Tris HCl pH 8,0, 200 mM NaCl, 5 mM EDTA, 0,05% SDS, 1,4 mg/mL Proteinasa K (NEB) y 12,5 mM DTT) durante 12-20 horas a 55 °C. Para extraer el ADNg, se combinó el lisado con 165 µL de perlas paramagnéticas, se mezcló bien, se incubó durante 5 minutos, se separó en una placa magnética y se lavó 3 veces con EtOH al 70%, se dejó secar durante 5 minutos y se eluyó en tampón de elución 65 µE (1,2 mM Tris-HCl pH 8,0). Los loci genómicos se amplificaron mediante PCR con Phusion Hot Start Flex DNA Polymerase (New England Biolabs; NEB) utilizando 100 ng de ADNg como plantilla y los cebadores enumerados en la Tabla 3. Tras el análisis en una máquina de electroforesis capilar QIAxcel (Qiagen), los productos de la PCR se purificaron con perlas paramagnéticas.

Para los experimentos con nucleasas, el porcentaje de modificación de los sitios diana humanos endógenos se determinó mediante ensayos de endonucleasa I T7 (T7EI), de forma similar a como se ha descrito antes (Reyon et al., Nat Biotechnol., 2012, 30:460-5). Brevemente, 200 ng de productos de PCR purificados fueron desnaturalizados,

hibridados y digeridos con 10 U de T7EI (NEB) a 37 °C durante 25 minutos. Los digeridos se purificaron con perlas paramagnéticas y se analizaron utilizando una máquina QIAxcel para estimar la modificación del sitio diana.

Para los experimentos de edición de bases, la secuenciación profunda dirigida se realizó esencialmente como se ha descrito antes (Kleinstiver et al., Nature, 2016, 529:490-5). Se generaron bibliotecas Tru-seq de doble índice a partir de productos de PCR purificados y agrupados utilizando un kit de preparación de bibliotecas KAPA HTP (KAPA BioSystems) y se secuenciaron en un secuenciador Illumina MiSeq. Las muestras se secuenciaron con un recuento medio de lectura de 55.000 y un mínimo de 8.500 lecturas. Las sustituciones de nucleótidos y las mutaciones por inserción o delección (indels) se analizaron utilizando una versión modificada de CRISPResso (Pinello et al., Nat Biotechnol., 2016, 34:695-7), realizando un análisis personalizado adicional para examinar los porcentajes de indels (definidos como [lecturas modificadas - lecturas solo de sustitución] / lecturas totales *100), en una ventana de 44 nt que abarcaba la región de -14 a +30 de cada sitio diana (10 nt adicionales en sentido ascendente de 4 nt PAM y 10 nt en sentido descendente de la secuencia espaciadora de 20 nt).

GUIDE-seq.

Los experimentos GUIDE-seq se realizaron como se ha descrito previamente (Tsai et al., Nat Biotechnol., 2015, 33:187-197). Resumiendo, las células U2OS se electroporaron como se ha descrito antes, pero incluyendo 100 pmol del marcador GUIDE-seq de oligodesoxinucleótidos de doble cadena (dsODN). Se realizaron ensayos de fragmentos de restricción de longitud polimórfica (RFLP) (como se ha descrito antes; Kleinstiver et al., Nature, 2015, 523:481-5) y ensayos T7E1 (como se ha descrito antes) para evaluar la integración del marcador GUIDE-seq y los porcentajes de modificación *on-target*, respectivamente. Las bibliotecas GUIDE-seq se secuenciaron utilizando un secuenciador Illumina MiSeq, y los datos se analizaron utilizando guideseq (Tsai et al., Nat Biotechnol., 2016, 34:483) v1.1 con una ventana de 75 pb y permitiendo hasta 9 discordancias antes del procesamiento de datos en sentido descendente. Se identificaron polimorfismos de un solo nucleótido (SNP) de alta confianza específicos del tipo celular, utilizando SAMTools.

Experimentos de activación génica.

Para los experimentos con ARNcr o ARNsg focalizados al promotor VEGFA, se sembraron $1,6 \times 10^{15}$ de células HEK293 por pocillo en placas de 24 pocillos aproximadamente 24 horas antes de la transfección con plásmidos que codificaban activadores Cas12a o Cas9 y grupos de ARNcr o ARNsg (750 ng y 250 ng, respectivamente) 1,5 μ L TransIT-LT1 (Mirus), y Opti-MEM a un volumen total de 50 μ L. Los medios de cultivo celular se cambiaron 22 horas después de la transfección, y se recogieron alícuotas del sobrenadante de los medios 44 horas después de la transfección para determinar la concentración de VEGFA utilizando un kit ELISA Quantikine de VEGF humano (R&D Systems).

Para los experimentos con ARNcr que seleccionan los promotores AR, HBB o NPY1R, se sembraron $8,6 \times 10^4$ de células HEK293 por pocillo en placas de 12 pocillos aproximadamente 24 horas antes de la transfección con 750 ng de plásmido de expresión del activador de Cas12a, 250 ng de grupos de plásmidos de ARNcr, 3 μ L TransIT-LT1 (Mirus), y 100 μ L Opti-MEM. El ARN total se extrajo de las células transfectadas 72 horas después de la transfección utilizando el kit NucleoSpin RNA Plus (Clontech). La síntesis de ADNc utilizando un kit de ARN a ADNc de alta capacidad (ThermoFisher) se realizó con 250 ng de ARN purificado, y 3 μ L de ADNc diluido a 1:20 se amplificaron mediante PCR cuantitativa de transcripción inversa (RT-qPCR) utilizando Fast SYBR Green Master Mix (ThermoFisher) y los cebadores enumerados en la Tabla 3. Las reacciones de RT-qPCR se realizaron en un LightCycler480 (Roche) con el siguiente programa de ciclos: desnaturalización inicial a 95 °C durante 20 segundos (s) seguida de 45 ciclos de 95 °C durante 3 s y 60 °C durante 30 s. Si la amplificación de la muestra no alcanza el umbral de detección tras 35 ciclos, los valores de Ct (ciclos hasta el umbral) se consideran 35 debido a las fluctuaciones de Ct típicas de las transcripciones expresadas a niveles muy bajos. Los niveles de expresión genética en los experimentos de controles negativos (activador de Cas12a y plásmidos de ARNcr vacíos) se normalizaron a la expresión de HPRT1.

Expresión y purificación de proteínas Cas12a.

Los plásmidos que codifican proteínas de fusión Cas12a-SV40NLS-6xHis se transformaron en Rosetta 2 (DE3) *E. coli*, y se inocularon colonias individuales en 25 ml de cultivos de medio LB que contenían 50 mg/L de kanamicina y 25 mg/L de cloranfenicol (Kan/Cm) antes del crecimiento a 25 °C durante 16 horas. Después, los cultivos iniciadores se diluyeron 1: 100 en 150 mL de medio LB conteniendo Kan/Cm y se cultivaron a 37 °C hasta que OD₆₀₀ alcanzó 0,4. A continuación, los cultivos se indujeron con 0,2 mM de isopropil β -D-tiogalactopiranosido antes de proceder al agitado a 18 °C durante 23 horas. Los pellets celulares de 50 mL del cultivo se recogieron por centrifugación a 1200 g durante 15 minutos y se suspendieron en 1 mL de tampón de lisis que contenía 20 mM Hepes pH 7,5, 100 mM KCl, 5 mM MgCl₂, 5% de glicerol, 1 mM DTT, inhibidor de proteasas Sigmafast (Sigma-Aldrich) y 0,1% de Triton X-100. La suspensión celular se cargó en un milliTUBE de fibra AFA de 1 mL (Covaris) y

se lisó utilizando un ultrasonificador focalizado E220evolution (Covaris) de acuerdo con las siguientes condiciones: potencia de intensidad pico de 150 W, 200 ciclos por ráfaga, factor de trabajo del 10% y tratamiento durante 20 minutos a 5 °C. El lisado celular se centrifugó durante 20 minutos a 21.000 g y 4 °C, y el sobrenadante se mezcló con un volumen igual de tampón de unión (tampón de lisis + 10 mM de imidazol), se añadió a 400 µL de resina HisPur Ni-NTA (Thermo Fisher Scientific) preequilibrada en tampón de unión, y se agitó a 4 °C durante 8 horas. La resina unida a proteínas se lavó tres veces con 1 mL de tampón de lavado (20 mM Hepes pH 7,5, 500 mM KCl, 5 mM MgCl₂, 5% de glicerol, 25 mM de imidazol y 0,1% de Triton X-100), se lavó una vez con 1 mL de tampón de unión y después se hicieron tres eluciones secuenciales con 500 µL de tampón de elución (20 mM Hepes pH 7,5, 100 mM KCl, 5 mM MgCl₂, 10% de glicerol y 500 mM de imidazol). Las eluciones seleccionadas se agruparon y dializaron utilizando tubos de diálisis de celulosa estándar Spectra/Por 4 (Spectrum Chemical Manufacturing Corp) en tres intercambios secuenciales de tampón 1:500, los dos primeros en tampón de diálisis (300 mM NaCl, 10 mM Tris-HCl pH 7,4, 0,1 mM EDTA y 1 mM DTT) y el último en tampón de diálisis con un 20% de glicerol. A continuación, las proteínas se concentraron con unidades de filtro centrífugo Amicon Ultra-0,5 ml (Millipore Sigma), se diluyeron con un volumen igual de tampón de diálisis con un 80% de glicerol y se almacenaron a -20 °C.

15 Reacciones de escisión *in vitro*.

Los ARNcr Cas12a se transcribieron *in vitro* a partir de aproximadamente 1 µg de plásmido de transcripción de ARNcr linealizado HindIII utilizando el kit T7 RiboMAX Express Large Scale RNA Production (Promega) a 37 °C durante 16 horas. La plantilla de ADN se degradó añadiendo 1 µL de DNasa RQ1 a 37 °C durante 15 minutos, y el ARN se purificó posteriormente con el mini kit miRNeasy (Qiagen). Las reacciones de escisión *in vitro* consistieron en 25 nM de plásmido de sustrato linealizado-Pvul, 300 nM de ARNcr y 200 nM de proteína Cas12a purificada en tampón de escisión (10 mM Hepes pH 7,5, 150 mM NaCl y 5 mM MgCl₂), y se realizaron a 37 °C salvo que se indicara lo contrario. Los sustratos de plásmidos para los ensayos de tolerancia a la temperatura codificaban el espaciador del sitio 2 de PAMDA con un PAM TTTA. Se prepararon las mezclas maestras de la reacción de escisión y se alicuotaron en volúmenes de 5 µL para cada punto temporal, se incubaron en un termociclador y se interrumpieron añadiendo 10 µL de tampón de parada (0,5% SDS, 50 mM EDTA). Los alicuotas detenidos se purificaron con perlas paramagnéticas y el porcentaje de escisión se cuantificó con el software QIAxcel ScreenGel (v1.4).

Ensayo de determinación de PAM.

Las bibliotecas de plásmidos que codifican sitios diana con secuencias aleatorias se clonaron utilizando Klenoq(-exo) (NEB) para rellenar en las cadenas inferiores de dos oligos separados que presentaban secuencias 5' aleatorias de 10 nt de dos secuencias espaciadoras distintas (Tabla 3). El producto de doble cadena se digirió con EcoRI y se ligó en EcoRI y SphI digerido p11-lacY- wtx1 (plásmido Addgene 69056; donación de Huimin Zhao). Los ligandos se transformaron en *E. coli* XL1 Blue electrocompetente, se recuperaron en 9 mL de SOC a 37°C durante 1 hora y después se cultivaron durante 16 horas en 150 mL de medio LB con 100 mg/L de carbenicilina. Se estimó que la complejidad de cada biblioteca era superior a 10⁶ basado en el número de transformantes observados.

Las reacciones de escisión de las bibliotecas de plásmidos de PAM aleatorizados se realizaron como se ha descrito antes, deteniendo los alicuotas a los 3, 6, 12, 24 y 48 minutos. Las reacciones se purificaron con perlas magnéticas y se utilizó aproximadamente 1-5 ng como plantilla para la amplificación de PCR de las moléculas no escindidas con Phusion Hot Start Flex DNA Polymerase (NEB) durante 15 ciclos. Durante las reacciones de PCR, se añadió un índice molecular único (UMI) de 4 nt en sentido ascendente de PAM para permitir el demultiplexado de las muestras de puntos temporales, y también se generaron productos a partir de un plásmido no digerido para determinar la representación inicial de PAM en las bibliotecas. Los productos de PCR purificados se cuantificaron con el sistema QuantiFluor dsDNA (Promega), se normalizaron y se agruparon para la preparación de bibliotecas con adaptadores de doble índice Illumina utilizando un Kit KAPA HTP PCR-free Library Preparation (KAPA BioSystems). Las bibliotecas se cuantificaron utilizando el Kit Universal KAPA Illumina Library qPCR Quantification (KAPA Biosystems) y se secuenciaron en un secuenciador Illumina MiSeq utilizando un kit de 300 ciclos v2 (Illumina).

Las lecturas de secuenciación se analizaron utilizando un script Python personalizado para estimar las tasas de escisión en cada PAM para una proteína determinada. Las lecturas emparejadas de extremos se filtraron por puntuación Phred ($\geq$ Q30) y después se fusionaron con el requisito de coincidencias perfectas de UMIs de puntos temporales, PAM y la secuencia espaciadora. Se generaron recuentos por cada PAM de 4 y 5 nt para todos los puntos temporales, proteínas y espaciadores. Los recuentos de PAM se corrigieron para tener en cuenta las diferencias entre muestras en la profundidad de secuenciación, se convirtieron en una fracción de la representación inicial de ese PAM en la biblioteca de plásmidos original (según lo determinado por el control no digerido) y se normalizaron para tener en cuenta el aumento de la representación fraccional de sustratos no escindidos con el tiempo debido al agotamiento de los sustratos escindidos (seleccionando los 5 PAMs con los recuentos medios más altos en todos los puntos temporales para representar el perfil de sustratos no escindibles). El agotamiento de

cada PAM a lo largo del tiempo se ajustó a un modelo de decaimiento exponencial ($y(t) = Ae^{(-kt)}$, donde $y(t)$ es el recuento normalizado de PAM, t es el tiempo (minutos), k es la constante de velocidad y A es una constante), mediante regresión lineal de mínimos cuadrados.

Cálculos del rango de selección.

- 5 Se evaluaron los rangos de selección de nucleasas AsCas 12a de tipo salvaje y variante en varios elementos genómicos anotados utilizando el archivo GTF de la versión 27 de GENCODE. Se enumeraron las apariciones completas de PAMs de 4 nt seleccionables en regiones que abarcaban 1 kb en sentido ascendente de todos los sitios de inicio de la transcripción (TSS), en el primer exón de todos los genes y en todos los ARNm anotados. Los valores de los parámetros para cada elemento en el archivo GTF fueron: Exón1, característica tipo exón, exón_número 1, gen_tipo proteína_codificante; TSS, característica tipo transcripción, gen_tipo proteína_codificación o ARNm; ARNm, característica tipo gen, gen_tipo ARNm. Para cada elemento, los recuentos de PAM se normalizaron por longitud y se visualizaron mediante un diagrama de cajas. El script de identificación y enumeración de PAM se pondrá a disposición a solicitud. Los PAMs seleccionables para las nucleasas Cas12a incluyen: TTTV, para AsCas12a de tipo salvaje; TTYN, RTTC, CTTV, TATM, CTCC, TCCC, TACA (nivel 1), y RTTS, TATA, TGTV, ANCC, CVCC, TGCC, GTCC, TTAC (nivel 2) PAMs para eAsCas12a (ver Fig. 1g y datos ampliados Fig. 5h); TATV, AsCas12a-RVR; y TYCV para AsCas12a-RR.

Ensayos de detección de ADN.

- Los complejos RNP de Cas12a-ARNcr se formaron incubando 500 nM de proteína de AsCas12a purificada y 750 nM de ARNcr sintetizado químicamente (IDT) a 4 °C durante 5 minutos. Todas las reacciones se realizaron en 10 mM Hepes pH 7,5, 150 mM NaCl, y 5 mM MgCl₂. Después, las RNP se diluyeron a 100 nM y se mezclaron con 250 nM de ADN plasmídico activador linealizado que presentaba un sitio diana coincidente con un PAM canónico (TTTA) o no canónico (CTTA o ACCT) o un sitio diana no relacionado (control negativo) en un volumen total de 15 µl. Las reacciones se dejaron a 37 °C durante 30 minutos, antes de incubar a 4 °C. Para la detección fluorescente, se mezclaron 10 µl de la reacción RNP/diana-ADN con un volumen igual de 100 nM de reporter fluorescente personalizado (IDT) en una placa negra de 384 pocillos de bajo volumen (Corning). Las reacciones de detección se realizaron a 37 °C durante tres horas con toma de las mediciones cada 60 segundos con $\lambda_{ex} = 485$ nm y $\lambda_{em} = 528$ nm, usando un Synergy HTX Microplate Reader (BTX).

Secuencias

En los siguientes ejemplos se utilizaron las siguientes construcciones.

- 30 **BPK3079:** U6-AsCas12a-crRNA-BsmBlcassette

Promotor U6 en negro, ARNcr AsCas12a en cursiva, guanina necesaria para la transcripción U6 en **negrita**, casete de entrada espaciador en minúsculas con los sitios BsmBI doble subrayado, terminador U6 **(TTTTTTT) doble subrayado en negrita**.

TGTACAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCCTTCATATTTGCATATACGATACAAGGC
TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACG
TGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAATTTATGTTTTAAATGG
ACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGG
AAAGGACGAAACACCGTAATTTCTACTCTTGTAGATgagagacgattaatgcgtctccTTTTTTT(SEQ
ID NO: 415)

- 35 **BPK3082:** U6-LbCas12a-crRNA-BsmBlcassette

Promotor U6 en verde, ARNcr LbCas12a coloreado en cursiva, guanina necesaria para la transcripción U6 en **negrita**, casete de entrada espaciador en minúsculas con los sitios BsmBI doble subrayado, terminador U6 **(TTTTTTT) doble subrayado en negrita**.

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
 CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC
 TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACG
 TGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGG
 ACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATCTTGTGG
 AAAGGACGAAACACCGAATTTCTACTAAGTGTAGATgagagacgattaatgcgtctccTTTTTTT

(SEQ ID NO: 416)

BPK4446: U6-FnCas12a-crRNA-BsmBcassette

Promotor U6 en negro, ARNcr FnCas12a coloreado en cursiva, guanina necesaria para la transcripción U6 en **negrita**, casete de entrada espaciador en minúsculas con los sitios BsmBI doble subrayado, terminador U6 **(TTTTTTT) doble subrayado en negrita**

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
 CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC
 TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACG
 TGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGG
 ACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATCTTGTGG
 AAAGGACGAAACACCGAATTTCTACTGTTGTAGATgagagacgattaatgcgtctccTTTTTTT (SEQ
 ID NO: 417)

BPK4449: U6-MbCas12a-crRNA-BsmBcassette

Promotor U6 en negro, ARNcr MbCas12a coloreado en cursiva, guanina necesaria para la transcripción U6 en **negrita**, casete de entrada espaciador en minúsculas con los sitios BsmBI doble subrayado, terminador U6 **(TTTTTTT) doble subrayado en negrita**

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
 CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC
 TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACG
 TGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGG
 ACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATCTTGTGG
 AAAGGACGAAACACCGAATTTCTACTGTTGTAGATgagagacgattaatgcgtctccTTTTTTT
 (SEQ ID NO: 418)

MSP3491: T7-AsCas12a-crRNA-Bsalcassette

Promotor T7 en negro, guanina necesaria para la transcripción T7 en **negrita**, ARNcr AsCas12a en cursiva, casete de entrada espaciador en minúsculas con sitios BsaI doble subrayado, sitios de restricción para Drr1 (tttaaa) en minúsculas y **negrita** y HindIII **(AAGCTT) doble subrayado en negrita** para linealización.

TAATACGACTCACTATAGTAATTTCTACTCTTGTAGATggagacccatgcatagcgttgcggaatag
aatTTTgaacagattcaccaacacctagtggctcctTTaaa**AAGCTT** (SEQ ID NO: 419)

MSP3495: T7-LbCas12a-crRNA-Bsalcassette

- 5 Promotor T7 en negro, guanina necesaria para la transcripción T7 en **negrita**, ARNcr LbCas12a en cursiva, casete de entrada espaciador en minúsculas con sitios BsaI doble subrayado, sitios de restricción para DraI (TTAAA) en minúsculas y **negrita** y HindIII (**AAGCTT**) doblemente subrayado en negrita para linealización:

TAATACGACTCACTATAGTAATTTCTACTAAGTGTAGATggagacccatgcatagcgttgcggaatag
aatTTTgaacagattcaccaacacctagtggctcctTTaaa**AAGCTT** (SEQ ID NO:420)

Secuencia de nucleótidos de pCAG-humanAsCpf1-NLS-3xHA

- 10 AsCpfI humano con codón optimizado en letra normal (NTs 1-3921), NLS en minúsculas (aaaaggccggcgccacgaaaag- gccggccaggcaaaaagaaaaag, SEQ ID NO:21), marcador 3xHA (TACCCATACGATGTTCCAGATTACGCTTATCCCTAC-GACGTGCCTGATTATGCATACCCAT ATGATGTCCCCGACTATGCC, SEQ ID NO:5) en **negrita**

ATGACACAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGT
TTGAGCTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAG

GAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTAC
 AAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGC
 CGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGG
 AGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGA
 CCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGT
 TTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCC
 CTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGA
 ACGTGTTACAGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTGCCAGGACAAC
 TCCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCGTGATCACCGCCGTGCCAGCC
 TGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCG
 AGGAGGTGTTTTCTTCCCTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTA
 TAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCCTGA
 ACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCACATCATCGCTCCC
 TGCCACACAGATTTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCTGTCTTT
 CATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCCTTCTGCAAGTACAAGAC
 ACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACA
 GCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCC
 TGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGA
 CAGGCAAGATCACCAAGTCTGCCAAGGAGAAAGGTGCAGCGCAGCCTGAAGCACGAGGAT
 ATCAACCTGCAGGAGATCATCTCTGCCGCAAGGAGCTGAGCGAGGCCCTCAAGCA
 GAAACCAGCGAGATCCTGTCCACGACACGCCGCCCTGGATCAGCCACTGCCTACAA
 CCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGC
 CTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTT
 CTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAA
 GGCCAGAAATTATGCCACCAAGAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCA
 GATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAGAACAATGGCGCCATCCT
 GTTTGTGAAGAACGGCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAA
 GGCCCTGAGCTTCGAGCCACAGAGAAAACCGCGAGGGCTTTGATAAGATGTACTATGA
 CTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGAC
 AGCCCACTTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTTCATCGAGCCTCTG
 GAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAG
 ACAGCCTACGCCAAGAAAACCGCGACCCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTG
 GATCGACTTCACAAGGGATTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCT
 AGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCC
 CTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAG
 ACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGC
 AAGCCTAATCTGCACACACTGTATTGGACCGCCTGTTTTCTCCAGAGAACCTGGCCAAG
 ACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAG
 AGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAAC
 CCCAATCCCCGACACCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCA

CGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTCTC
 ACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCAC
 ACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCT
 GAAGGAGCACCOCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCT
 ATATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCC
 AGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGG
 CAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGT
 CATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGGTGGAGAACC
 TGAATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAG
 TTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGGTGAAGGACTATCCAGCAGAG
 AAAGTGGGAGGGCTGCTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCAAG
 ATGGGCACCCAGTCTGGCTTCTGTCTTTACGTGCCTGCCCATATACATCTAAGATCGATC
 CCCTGACCGGCTTCTGTGGACCCCTTCTGTGGAAAACCATCAAGAATCACGAGAGCCGCA
 AGCACTTCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCC
 TGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCCCTGCCCGGCTTTATGCCTG
 CATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCTTTCA
 TCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGG
 GACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAG
 GGATGGCTCCAAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACAC
 CATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCG
 AGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTC
 AGAACCAGAGTGGCCCATGGACGCGGATGCCAATGGCGCCTACCACATCGCCCTGAAG
 GGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCAT
 CTCCAATCAGGACTGGCTGGCTTACATCCAGGAGCTGCGCAAC**aaaaggccggccacgaaaa**
aggccggccaggcaaaaaagaaaaagGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCT
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID
 NO:6)

Secuencia de aminoácidos de AsCpf1-NLS-3xHA

AsCpf1 en letra normal (AAs 1-1306), NLS (krpaatkkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA(YPY-DVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQGFIEEDKARNNDHYKELKPIIDRIYKT
YADQCLQLVQLDWNLSAAIDSYRKEKTEETRNLIEEQATYRNAIHDIYFIGRTDNLTDANKR
HAEIYKGLFKAELFNGKVLKQLGTVTTEHENALLRSFDKFTTYFSGFYENRKNVFS AEDISTAI
PHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQID
LYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFK
SDEEVQSFSCKYKTLNRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALY
ERRISELTGKITSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAALDQPLPT
TLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNY
ATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT

EKTSEGFDMYDYFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNPNPE
KEPKKFQYAYAKKTGDQKGYREALCKWDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYA
ELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHGKPNLHTLYWTGLFSPENLAKT
SIKLNQGAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIDTLYQELYDYVNHRLSHDLSD
EARALLPNVITKEVSHEIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGI
DRGERNLIYITVIDSTGKILEQRSNTIQQFDYQKKLDNREKERVAAARQAWSVVGTIKDLKQGYL
SQVIHEIVDLMIHYQAVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVG
GVLNPLYQLTDQFTSFAMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLG
FDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFQDAKGTPIAGKRIVPVI
ENHRFTGRYRDLYPANELIALLEEKGIIVFRDGSNILPKLLENDSDHAIDTMVALIRSVLQMRNSN
AATGEDYINSPVRDLNGVCFDSRFQNPWPMDADANGAYHIALKGQLLLNHLKESKDLKLQN
GISNQDWLAYIQELRNkrpaafkkagqakkkkGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ
ID NO:9)

SQT1659: pCAG-hAsCas12a-NLS-3xHA

Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, nucleoplasmina NLS
(aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias
5 enlazadoras *en cursiva*, marcador 3xHA

(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTC-
CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGT
 TTGAGCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGG
 AGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACA
 AGACCTATGCCGACCAAGTGCTGCAAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCG
 CCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAG
 CAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCCGACAGACAACCTGACC
 GATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCTGTTCAGGCCGAGCTGTIT
 AATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCAACAACCGAGCACGAGAACGCCCT
 GCTGCCGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGAA
 CGTGTTGAGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTGCAAGGACAACCTT
 CCCCAGTTTAAAGGAGAATGTGCACATCTTCACACGCCCTGATCACCGCCGTGCCACGCCCTG
 CCGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAG
 GAGGTGTTTTCTTCCCTTTTATAACCAAGCTGCTGACACAGACCCAGATCGACCTGTATA
 ACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCCTGAAC
 GAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCCCTG
 CCACACAGATTATCCCCCTGTTTAAAGCAGATCCTGTCCGATAGGAACACCCCTGTCTTTCA
 TCCTCGAAGGAGTTTAAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACAC
 TGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAAAGAGCTGAACAGC
 ATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTG

TGGGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACA
 GGCAAGATCACCAGTCTGCCAAGGAGAAAGTGCAGCGCAGCCTGAAGCACCAGGATAT
 CAACCTGCAGGAGATCATCTCTGCCGAGGCAAGGAGCTGAGCGAGGCTTCAAGCAGAA
 AACCAGCGAGATCCTGTCCACGCACACGCCGCCCTGGATCAGCCACTGGCTACAACCCCT
 GAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCCTGT
 ACCACCTGCTGGACTGGTTTGGCGTGGATGAGTCCAAACGAGGTGGACCCCGAGTTCCTGQ
 CCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAG
 AAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAGCTGAACCTTTCAGATGCC
 TACACTGGCTCTTGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCTGTGTTGT
 GAAGAACGGCTGTACTATCTGGGCATCATGCCAAAGCAGAAAGGCGAGGTATAAGGCCCT
 GAGCTTCGAGCCACAGAGAAAACCGAGCGAGGCTTTGATAAGATGTACTATGACTACTT
 CCTGTATGCCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCA
 CTTCAGACCCACACAACCCCATCTGCTGTCTCAACAATTTTCATCGAGCCTCTGGAGATC
 ACAAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGC
 CTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGGCCCTGTGCAAGTGGATCG
 ACTTCACAAGGGATTTTCTGTCCAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCT
 GCGGCCATCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTTGCT
 GTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAG
 GCAAGCTGTACCTGTTCAGATCTATAACAAGGACTTTGCCAAGGGCCACCCACGGCAAGC
 CTAATCTGCACACACTGTATTGGACCGGCTGTCTTCTCCAGAGAACCTGGCCAAAGACAAG
 CATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGAT
 GGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAA
 TCCCGACACCTGTACAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCT
 GTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCAGGAT
 CATCAAGGATAGGGCTTTACCAAGGACAAGTCTTTTCCACGTGCCATCACACTGAAC
 TATCAGGCCGCCAATTCCCATCTAAGTTCAACCAGAGGGTGAATGCCCTACCTGAAGGAG
 CACCCCGAGACACCTATCATCGGCATCGATCGGGCGAGAGAACTGATCTATATLACA
 GTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTT
 GATTACCAGAAGAAGCTGGACAACAGGGAGAAAGGAGAGGGTGGCAGCAAGGCAGGCTG
 GTCTGTGGTGGGACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTATCCACGA
 GATCGTGGACCTGATGATCCACTACCAAGCCGTGGTGGTGTGGGAGAACCCTGAATTTCGG
 CTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGA
 TGCTGATCGATAAGCTGAATTGCCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAG
 GCGTGTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCAAGATGGGCACCC
 AGTCTGGCTTCTGTTTTACGTGCTGCCCATATACATCTAAGATCGATCCCTGACCGG
 CTTCGTGGACCCCTTCGTGTGGAACCATCAAGAATCAGAGAGCCGCAAGCAGCTTCCT
 GGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCTGCACTTTAAG
 ATGAACAGAAATCTGTCCTCCAGAGGGGCTGCCCGGCTTTATGCTGCGATGGGATATCG
 TGTTCGAGAAGAAGCAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGA
 GAATCGTGCCAGTGATCGAGAATCACAGATTACCGGCAGATACCGGACCTGTATCTGT

CCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACA
TCTTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACCATGGTGGCCCTGA
TCCGCAGCGTGTGTCAGATGCGGAACCTCAATGCGGCCACAGGCGAGGACTATATCAACA
GCCCGGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCAGAGTGGCC
CATGGAAGCGCGATGCCAATGGCGCTACACATCGCCCTGAAGGGCCAGCTGCTGCTGAA
TCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCT
GGCCTACATCCAGGAGCTGCGCAACaaagggcgccggccacgaanagggcgccaggcaaaaagaanaagGUA
TCTTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCAT
ACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID NO: 421)

AAS826: pCAG-hAsCas12a(E174R/S542R)-NLS-3xHA

5 Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados (E174R/S542R) doble subrayado, nucleoplasmina NLS (aaaaggccggcgccacgaaaaaggccggccaggcaaaaa-gaaaaag, SEQ ID NO:21) en minúsculas, secuencias enlazadoras en *cursiva*, marcador 3xHA tag (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCA TATGATGTC-CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGA
GCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAG
GACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAA
GACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCC
GCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGG
AGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTG
ACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCT
GTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAG
GAAGAACGTGTTACGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTGCAG
GACAACCTCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCCTGATCACCGCCGTG
CCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCA
CCTCCATCGAGGAGGTGTTTTCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGA
TCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGAT
CAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCAC
ATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTAAGCAGATCCTGTCCGATAGG
AACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCTT
CTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTG
TTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGA
GACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAG
CGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGC
GCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGA
GCTGAGCGAGGCCCTCAAGCAGAAAACAGCGAGATCCTGTCCCACGCACACGCCGCC

CTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGT
 CTCAGCTGGACAGCCTGCTGGGCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGA
 GTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATG
 GAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTC
 CGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCAGGCTGGGACCTG
 AATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGG
 CATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAA
 ACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGAT
 CCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTAGAGCCACACAACC
 CCCATCCTGCTGTCCAACAATTCATCGAGCCTCTGGAGATCACAAGGAGATCTACGA
 CCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTACAGACAGCCTACGCCAAGAAAACCG
 GCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTT
 TCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGGGGCCATCCTCTCA
 GTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCT
 TCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCT
 GTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACA
 CACTGTATTGGACCGGCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTG
 AATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACC
 GGCTGGGAGAGAAGATGCTGAACAAGAAAGCTGAAGGATCAGAAAACCCCAATCCCGA
 CACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCTG
 ATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAGGAGGTGTCTCACGAGATCAT
 CAAGSATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTA
 TCAGGCCGCCAATTCCCCTCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGC
 ACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACA
 GTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGT
 TTGATTACCAGAAGAAGCTGGACAACAGGGAGAAAGGAGAGGGTGGCAGCAAGGCAGGC
 CTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCAATCC
 ACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTCTGGAGAACCTGAA
 TTTCCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTT
 GAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTCTGAAGGACTATCCAGCAGAGAA
 AGTGGGAGGGCTGCTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCAAGA
 TGGGCACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCATATACATCTAAGATCGATC
 CCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCG
 CAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCA
 TCCTGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATG
 CCTGCATGGGATATCGTGTTCGAGAAGAAGAGACACAGTTTGACGCCAAGGGCACCC
 CTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTACCCGGCAGA
 TACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCG
 TGTTCAAGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCC
 ATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCCGAACCTCCAATGCCG
 CCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGA
 CTCCCGGTTTCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCCTACCAC
 ATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGC
 TGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaaa
 agggcggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGA**TCCTACCCATACGATGTTCC**
AGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGGACT
ATGCCTAA (SEQ ID NO: 422)

Codón humano optimizado Acidaminococcus sp. BV3L6Cas12a (AsCas12a) en negro, codones modificados para eAsCas12a (E174R/S542R/K548R) doble subrayado, nucleoplasmina NLS

(aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias
 5 enlazadoras en *cursiva*, marcador 3xHA
 (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCA TATGATGTC-
 CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGA
 GCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAG
 GACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAA
 GACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCC
 GCCATCGAATCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGG
 AGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGACAGACAACCTG
 ACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCT
 GTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCAGGAGAAC
 GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGG
 AAGAACGTGTTGAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGG
 ACAACTTCCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCTGATCACCGCCGTGC
 CCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTGGTGAACAC
 CTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCAGCTGCTGACACAGACCCAGAT
 CGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATC
 AAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACAT
 CATCGCCTCCCTGCCACACAGATTATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGA
 ACACCTGTCTTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCTCTC
 TGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGT
 TTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAG
 ACAATCAGCAGCGCCCTGTGOGACCACTGGGATACACTGAGGAATGCCCTGTATGAGC
 GGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCG
 CAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAG
 CTGAGCGAGGCCTTCAAGCAGAAAACCGAGCGAGATCTGTCCACGCACACGCCGCC

TGGATCAGCCACTGCCTACAACCCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCT
 CAGCTGGACAGCCTGCTGGGCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGT
 CCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGA
 GCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGT
 GGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGCCCGAGGGCTGGGACGTGAATG
 GAGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATC
 ATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAAACCA
 GCGAGGGCTTTGATAAGATGTACTATGACTACTTCCTGATGCCGCCAAGATGATCCCA
 AAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTAGACCCACACAACCCCAT
 CCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGA
 ACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGGCAG
 CAGAAGGGCTACAGAGAGGGCCTGTGCAAGTGGATCGACTTCACAAGGATTTTCTGTG
 CAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAA
 GGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAGCTTCAGAG
 GAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCA
 GATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGT
 ATTGGACCGGCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGC
 CAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGG
 GAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCGACACCCGTG
 TACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGC
 CAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATA
 GCGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAACTATCAGGCCG
 CCAATTCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAG
 ACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCAGTGTATCGA
 CTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACC
 AGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTG
 TGGTGGGCACAAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTATCCACGAGATC
 GTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGTGGAGAACCTGAATTTCTGGCTT
 TAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCCAGAAGATG
 CTGATCGATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAGG
 CGTGTGAACCCATACCAGCTGACAGACCAGTTCACTCTTTGCCAAGATGGGCACCC
 AGTCTGGCTTCTGTTTTACGTGCCTGCCCATATACATCTAAGATCGATCCCTGACCG
 GCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTC
 CTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACCTT
 TAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGG
 ATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCC
 GGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACC
 TGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGA
 TGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACCA
 TGGTGGCCCTGATCCGACGCTGCTGCAGATGCGGAACCTCAATGCCGCCACAGGCGA
 GGAATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTC
 AGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAA
 GGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGC
 ATCTCCAATCAGGACTGGCTGGCTACATCCAGGAGCTGCGCAACaaaaggccggccgac
 gaaaaaggccggccaggcaaaaaaggaagGGATCCTACCCATACGATGTTCCAGATTACGCTT
 ATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA
 (SEQ ID NO:423)

AAS1815: pCAG-heAsCas12a-HF1 (E174R/N282A/S542R/K548R)-NLS-3xHA

Codón humano optimizado *Acidaminococcus* sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados (E174R/N282A/S542R/K548R) doblemente subrayados, nucleoplasmina NLS (aaaaggccggcggccacgaaaaaggccggccag-gcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias enlazadoras en *cursiva*, marcador 3xHA (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCA TATGATGTC-CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**

ATGACACAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGA
GCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAG
GACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAA
GACCTATGCCGACCACTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCC
GCCATCGACTCCTATAGAAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGG
AGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGACAGACAACCTG
ACCGATGCCATCAATAAGAGACAAGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCT
GTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGG
AAGAACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCCGATCGTGCAGG
ACAACCTCCCAAGTTTAAAGGAGAATTGTACATCTTCACACGCCCTGATCACCGCCGTGC
CCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCAC
CTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCAGCTGCTGACACAGACCCAGAT
CGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATC
AAGGGCCTGAACGAGGTGCTGCCCTCTGGCCATCCAGAAGAATGATGAGACAGCCACA
TCATCGCCTCCCTGCCACACAGATTATCCCCCTGTTAAGCAGATCCTGTCCGATAGGA
ACACCCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTC
TGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGT
TTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAG
ACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGC
GGAGAATCTCCGAGCTGACAGGCAAGATCACCAGCTGCCAAGGAGAAGGTGCAGCG
CAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAAGGCAAGGAG
CTGAGCGAGGCCCTTCAAGCAGAAAACAGCGAGATCCTGTCCACGCACACGCCGCC
TGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCT
CAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGT

CCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGA
 GCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGT
 GGAGAAGTTCAAGCTGAACCTTTCAGATGCCTACACTGGCCAGAGGCTGGGACGTGAAT
 GAGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATC
 ATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAAACCA
 GCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCA
 AAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCAT
 CCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAGGAGATCTACGACCTGA
 ACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGAC
 CAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTC
 CAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAA
 GGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAGCTTCCAGA
 GAATGCCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCCAAGCTGTACCTGTTCCA
 GATCTATAACAAGGACTTTGCCAAGGGCCACCAAGGCAAGCCTAATCTGCACACACTGT
 ATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGC
 CAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGG
 GAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCCTG
 TACCAGGAGCTGTACGACTATGTGAATCAGACTGTCCACGACCTGTCTGATGAGGC
 CAGGGCCCTGCTGCCAACGATGATCACCAGGAGGTGTCTCAGAGATCATCAAGGATA
 GGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAAGTATCAGGCCG
 CCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCCTACCTGAAGGAGCACCCCGAG
 ACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCAGTGATCGA
 CTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACC
 AGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTG
 TGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCAGAGATC
 GTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGTCTGGAGAACCTGAATTTGGCTT
 TAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCCAGAAGATG
 CTGATCGATAAGCTGAATTGCCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAGG
 CGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCC
 AGTCTGGCTTCTGTTTTACGTGCCTGCCCATATACATCTAAGATCGATCCCTGACCG
 GCTTCGTGGACCCCTTCGTGTGGAAGCCATCAAGAATCAGGAGCCGCAAGCACTTC
 CTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACCTT
 TAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGG
 ATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTTCATCGCC
 GGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACC
 TGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGA
 TGGCTCCAACATCCTGCCAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACCA
 TGGTGGCCCTGATCCGAGCGTCTGCAGATGCCGAACCTCAATGCCGCCACAGGCGA
 GGAATATCAACAGCCCGTGCAGATCTGAATGGCGTGTCTTCGACTCCCGGTTTC
 AGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCTACACATCGCCCTGAA
 GGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGC
 ATCTCCAATCAGGACTGGCTGGCTACATCCAGGAGCTGCGCAACaaaaggccggccgac
 gaaaaaggccggccaggccaaaaaagaaaaagGGA TCCTACCCATACGATGTTCCAGATTACGCT
 TATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA
 (SEQ ID NO: 424)

BPK3541: pET-28b-hAsCas12a-NLS-6xHis

Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones con mutaciones silenciosas para eliminar sitios NcoI doble subrayado, glicina insertada subrayado punteado, nucleoplasmina NLS (aaaaggccggccggccac- gaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias espaciadoras en cursiva, 6xHis en **negrita**

5

ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTT
TGAGCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAG
GAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTA
CAAGACCTATGCCGACCAGTGCCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGC
GCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCG
AGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAAC
CTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGA
GCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAG
AACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAAC
AGGAAGAACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCCGATCGTGC
AGGACAACCTTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATACCGCC
GTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGA
GCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCAGCTGCTGACACAGACCC
AGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGAGGCAGGCACCGAGAA
GATCAAGGGCCTGAACGAGGTGCTGAATCTGCCATCCAGAAGAATGATGAGACAGCC
CACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGAT
AGGAACACCCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTC
CTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCC
CTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCT
GGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTAT
GAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGC
AGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCCAA
GGAGCTGAGCGAGGCCTTCAAGCAGAAAACCGCGAGATCCTGTCCACGCACACGCC
GCCCTGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACTGCTGGACTGGTTTGCCGTGGA
TGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAG
ATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTA

CTCCGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCTCTGGCTGGGACG
TGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTG
GGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGA
AAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATG
ATCCCAAAGTGACGACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAAC
CCCCATCCTGCTGTCCAACAATTTGATCGAGCCTCTGGAGATCACAAAGGAGATCTACG
ACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACC
GGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATT
TTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTC
AGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGTGTACACATCAGC
TTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACC
TGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCAC
ACACTGTATTGGACCGGCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCT
GAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACAC
CGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCG
ACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCT
GATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCA
TCAAGGATAGGCGCTTTACCAGCGACAAGTCTTTTTCCACGTGCCTATCACACTGAAC
ATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAG
CACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCAC
AGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAG
TTTGATTACAGAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCGAGG
CCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTGATC
CACGAGATCGTGACCTGATGATCCACTACCAGGCCGTGGTGGTGTGGAGAACCTGA
ATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTT
CGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGA
AAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCAAG
ATGGGCACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCATATACATCTAAGATCGAT
CCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCC
GCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGCGCACTTC
ATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCTGCCCGGCTTTAT
GCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
CCTTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
GTGTTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGC
CATCGACACGATGGTGGCCCTGATCCGCAGCGTGTGACAGATGCGGAATCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCATCTGAATGGCGTGTGCTTCG
ACTCCCGGTTTCAGAACCCAGAGTGGCCCAATGGACGCCGATGCCAATGGCGCTACCA
CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaa

aaggccggccgccaagaaaaaggccggccaggcaaaaaagaaaaagGGAGCGGCCGCACTCGAGCACC
ACCACCACCACCACTGA (SEQ ID NO:425)

RTW645: pET-28b-bLbCas12a-NLS-6xHis

- 5 Codón bacteriano optimizado Bacteria Lachnospiraceae ND2006 Cas12a (LbCas12a) en negro, glicina insertada subrayado punteado, NLS de nucleoplasmina (aaaaggccggccgccaagaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias enlazadoras en cursiva, 6xHis en **negrita**

ATGGGGAGCAAACCTGGAAAAATTTACGAATTGTTATAGCCTGTCCAAGACCCTGCGTTTC
 AAAGCCATCCCCGTTGGCAAAACCCAGGAGAATATTGATAATAAACGTCTGCTGGTTGA
 GGATGAAAAAAGAGCAGAAGACTATAAGGGAAGTCAAAAAACTGCTGGATCGGTACTACC
 TGAGCTTTATAAATGACGTGCTGCATAGCATTAACTGAAAAATCTGAATAACTATATTAG
 TCTGTTCCGCAAGAAAACCCGAACAGAGAAAGAAAAATAAGAGCTGGAAAACCTGGAGA
 TCAATCTGCGTAAAGAGATCGCAAAAGCTTTTAAAGGAAATGAAGGTTATAAAAGCCTGT
 TCAAAAAAGACATTATTGAAACCATCCTGCCGGAATTTCTGGATGATAAAGACGAGATAG
 CGCTCGTGAACAGCTTCAACGGGTTACGACCGCCTTCACGGGCTTTTTCGATAACAGG
 GAAAAATATGTTTTAGAGGAAGCCAAAAGCACCTCGATAGCGTTCCGTTGCATTAAATGAA
 AATTTGACAAGATATATCAGCAACATGGATATTTTCGAGAAAGTTGATGCGATCTTTGACA
 AACATGAAGTGCAGGAGATTAAGGAAAAATTTCTGAACAGCGATTATGATGTTGAGGATT
 TTTTCGAGGGGGAATTTTTAACTTTGTACTGACACAGGAAGGTATAGATGTGTATAATG
 CTATTATCGGCGGGTTCTGTTACCGAATCCGCGAGAAAAATTAAGGGTCTGAATGAGTAC
 ATCAATCTGTATAACCAAAAGACCAAAACAGAACTGCCAAAATTCAAACCGCTGTACAAG
 CAAGTCTGAGCGATCGGGAAAGCTTGAGCTTTTACGGTGAAGGTTATACCAGCGACGA
 GGAGGTACTGGAGGTCTTTTCGCAATACCCCTGAACAAGAACAGCGAAATTTTCAGCTCCA
 TTA AAAAGCTGGAGAACTGTTTAAAGATTTTGACGAGTACAGCAGCGCAGGTATTTTTG
 TGAAGAACGGACCTGCCATAAGCACCATTAAGCAAGGATATTTTTGGAGAGTGAATGTTA
 TCCGTGATAAATGGAACGCGGAATATGATGACATACACCTGAAAAAGAAGGCTGTGGTA
 ACTGAGAAATATGAAGACGATCGCCGCAAAAGCTTTAAAAAAATCGGCAGCTTTAGCCTG
 GAGCAGCTGCAGGAATATGCGGACGCCGACCTGAGCGTGGTCGAGAACTGAAGGAAA
 TTATTATCCAAAAAGTGGATGAGATTTACAAGGTATATGGTAGCAGCGAAAAACTGTTTG
 ATGCGGACTTCGTTCTGAAAAAAGCCTGAAAAAAATGATGCTGTTGTTGCGATCATGA
 AAGACCTGCTCGATAGCGTTAAGAGCTTTGAAAATTACATTAAAGCATTCTTTGGCGAGG
 GCAAAGAAACAAACAGAGACGAAAGCTTTTATGGCGACTTCGTCCTGGCTTATGACATCC
 TGTTGAAGGTAGATCATATATATGATGCAATTCGTAATTACGTAACCCAAAAGCGGTACA
 GCAAAGATAAGTTCAACTGTATTTCCAGAACCCGAGTTTATGGGTGGCTGGGACAAA
 GACAAGGAGACAGACTATCGCGCCACTATTCTGCGTTACGGCAGCAAGTACTATCTCGC
 CATCATGGACAAAAAATATGCAAAGTGCTGCAGAAAAATCGATAAAGACGACGTGAACG
 GAAATTACGAAAAAGATTAATTATAAGCTGCTGCCAGGGCCCAACAAGATGTTACCGAAAG
 TATTTTTTCCAAAAATGGATGGCATACTATAACCCGAGCGAGGATATACAGAAGATTTA

CAAAAATGGGACCTTCAAAAAGGGGGATATGTTCAATCTGAATGACTGCCACAAACTGAT
CGATTTTTTTAAAGATAGCATCAGCCGTTATCCTAAATGGTCAAACGCGTATGATTTTAAT
TTCTCCGAAACGGAGAAATATAAGACATTGCTGGTTTCTATCGCGAAGTCGAAGAACAG
GGTTATAAAGTTAGCTTTGAATCGGCCAGCAAGAAAGAGGTTGATAAACTGGTGGAGGA
GGGTAAGCTGTATATGTTTCAGATTATAACAAAGACTTTAGCGACAAAAGCCACGGTAC
TCCTAATCTGCATACGATGTACTTTAAACTGCTGTTTGATGAGAATAACCACGGCCAAAT
CCGTCTCTCCGGTGGAGCAGAACCTTTTATGCGGCGTGCAGCCCTAAAAAGGAAGAAC
TGGTGGTGATCCCGCCAACAGCCCGATTGCTAACAAAAATCCAGATAATCCTAAGAAG
ACCACCACACTGTCTGATGATGTCTATAAGGATAAACGTTTCTCGGAAGACCAGTATGAA
TTGCATATACCGATAGCAATTAATAATGCCCAAAAACATTTTCAAAATCAACACTGAAG
TTCGTGTGCTGCTGAAACATGATGATAATCCGTATGTGATCGGAATTGACCGTGGGGAG
AGAAATCTGCTGTATATTGTAGTCGTTGATGGCAAGGGCAACATCGTTGAGCAGTATAGC
CTGAATGAAATAATTAATAATTTAACGGTATACGTATTAACCCGACTATCATAGCCCTGC
TGGATAAAAAGGAGAAAGAGCGTTTTGAGGCACGCCAAAATTGGACGAGCATCGAAAAAC
ATCAAGCAACTGAAGGCAGGATATATCAGCCAAGTAGTCCATAAAATCTGTGAAGTGGTG
GAGAAGTACGACGCTGTCAATTGCCCTGGAAGACCTCAATAGCGGCTTTAAAAACAGCCG
GGTGAAGGTGGAGAAACAGGTATACCAAAAGTTTGAAGAGATGCTCATTGATAAGCTGAA
CTATATGGTTGATAAAAAGAGCAACCCGTGCCCACTGGCGGTGCACTGAAAGGGTACC
AAATTACCAATAAATTTGAAAGCTTTAAAGCATGAGCACGCAGAAATGGGTTTATTTTTTA
TATACCAGCATGGCTGACGAGCAAGATTGACCCAGCACTGGTTTTGTCAATCTGCTGA
AAACCAATACACAAGCATTGCGGATAGCAAAAAATTTATTTGAGCTTCGACCGTATTAT
GTATGTTCCGGAGGAAGATCTGTTGAATTTGCCCTGGATTATAAAAACTTCAGCCGCAC
CGATGCAGATTATCAAAAAATGGAAGCTGTACAGTTATGGTAATCGTATACGTATCTTC
CGTAATCCGAAGAAAAACAATGTGTTGATTGGGAAGAGGTCTGTCTGACCAGCGCGTA
TAAAGAACTGTTCAACAAGTACGGAATAAATTATCAGCAAGGTGACATTCCGCGCACTGCT
GTGTGAACAGTCAGATAAAGCATTATATAGCAGCTTTATGGCGCTGATGAGCCTGATGCT
CCAGATCCGCAACAGCATAACCGGTGGCACAGATGTTGACTTTCTGATCAGCCCTGTGA
AGAATAGCGACGGCATCTTCTACGATTCCAGGAAGTATGAAGCACAGGAAAACGCTATT
CTGCCTAAAAATGCCGATGCCAACGGCGCCTATAATATTGCACGGAAGGTTCTGTGGGC
GATTGGACAGTTCAAGAAAGCGGAAGATGAGAAGCTGGATAAGGTAAAAATTGCTATTA
GCAATAAGGAATGGCTGGAGTACGCACAGACATCGGTTAAACACGGTAGTaaaaggccggcg
gccacgaaaaaggccggccaggcaaaaaagaaaaagGGAGCGGCCGCACTCGAGCACCACCACCAC
CACCCTGA (SEQ ID NO:426)

AAS1885: pET-28b-heAsCas12a(E174R/S542R/K548R)-NLS-6xHis

- 5 Codones humanos optimizados para *Acidaminococcus* sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para eAsCas12a (E174R/S542R/K548R) en minúsculas y doble subrayado, codones con mutaciones silenciosas para eliminar los sitios NcoI en MAYÚSCULAS y doble subrayado, glicina insertada subrayado punteado, NLS de nucleoplasmina (aaaaggccggccggccacgaaaaag- gccggccaggcaaaaaagaaaaag, SEQ ID NO: 21) en minúsculas, secuencias enlazadoras en *cursiva*, 6xHis en **negrita**

ATGGGGACACAGTTGAGGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTG
 CGGTTTGAGCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCA
 TCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCG
 GATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAAC
 CTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCC
 TGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGAC
 GACAACCTGACCGATGCCATCAATAAGAGACAGGCCGAGATCTACAAGGCCCTGTTCAA
 GGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCCACAACCGAG
 CACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTAT
aggAACAGGAAGAACGTGTTGAGCGCCGAGGATATCAGCACAGCCATCCACACCGCAT
 CGTGACAGGACAACCTCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCAC
 CGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATOTTC
 GTGAGCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTTTATAACCAGCTGCTGACACAG
 ACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCG
 AGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACA
 GCCCACATCATCGCTCCCTGCCACACAGATTATCCCCCTGTTTAAGCAGATCCTGTC
 CGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCC
 AGTCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAG
 GCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAA
 GCTGGAGACAATCAGCAGCGCCCTGTCCGACCACTGGGATACACTGAGGAATGCCCTG
 TATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGG
 TGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGG
 CAAGGAGCTGAGCGAGGCCCTCAAGCAGAAAACCGAGATCCTGTCCACGACAC
 GCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCC
 TGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGT
 GGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTG
 GAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCC
 CTACTCCGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCaggGGCTGGG
 ACGTGAATaggGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATC
 TGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGGCCACAGA
 GAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGA
 TGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCCTTTTCAGACCCACACA
 ACCCCCATCCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTAC
 GACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAAC
 CGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGAT
 TTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCT
 CAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAG
 CTTCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTAC

CTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCA
CACACTGTATTGGACCGGCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGC
TGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACA
CCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCC
GACACCTGTACGAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTC
TGATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTCTCACGAGATC
ATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAAC
TATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGA
GCACCCCGAGACACCTATCATCGGCATCGATCGGGCGAGAGAAACCTGATCTATATCA
CAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCA
GTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAG
GCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCA
TCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGGTGGAGAACCT
GAATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAG
TTCCAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGGTGGTGGTGGTGGTGGTGGT
GAAAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCA
AGATGGGCACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCCATATACATCTAAGATCG
ATCCCCGTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGC
CGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTT
CATCCTGCACTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTAT
GCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
CCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTACCGGCAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
GTGTTCCAGGGATGGCTCCAACATCCTGCCAAGCTGCTGGAGAATGACGATTCTCACGC
CATCGACACGATGGTGGCCCTGATCCGCAGCGTGTGCGAGATGCGGAATCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTGG
ACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACGCCGATGCCAATGGCGCCTACCA
CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGGCAACaa
aaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaagGGAGCGGCCGCACTCGAGCACC
ACCACCACCACCACTGA SEQ ID NO:427)

AAS1880: pET-28b-hAsCas12a(E174R/S542R)-NLS-6xHis

Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados (E174R/S542R) en minúsculas doble subrayado, codones con mutaciones silenciosas para eliminar los sitios NcoI en MAYÚSCULAS y doble subrayado, glicina insertada subrayado punteado, NLS de nucleoplasmina (aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias enlazadoras en *cursiva*, 6xHis en **negrita**

5

ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTG
CGGTTTGAGCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCA
TCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCG
GATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAAC
CTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCC
TGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACA
GACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGCCTGTTCAA
GGCCGAGCTGTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAG
CACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTAT
ggAACAGGAAGAACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCGCAT
CGTGCAGGACAACCTCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCTGATCAC
CGCCGTGCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTC
GTGAGCACCTCCATCGAGGAGGTGTTTTCTTCCTTTTATAACCAGCTGCTGACACAG
ACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCCCG
AGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGATGATGAGACA
GCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTC
CGATAGGAACACCCTGTCTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCC
AGTCCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAG
GCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAA
GCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTG
TATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAGTCTGCCAAGGAGAAGG
TGCAGCGCAGCCTGAAGCAGGAGGATATCAACCTGCAGGAGATCATCTCTCCCGCAGG
CAAGGAGCTGAGCGAGGCCCTCAAGCAGAAAACCGAGGAGATCCTGTCCACGCACAC
GCCGCCCTGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCC
TGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGT
GGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTG
GAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCC
CTACTCCGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCggGGCTGGG
ACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTAT
CTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAG
AGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAG
ATGATCCCAAAGTGACAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTTCAGACCCACAC
AACCCCCATCCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAGGAGATCTA
CGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAA
CCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGA
TTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTC
TCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCA
GCTTCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTA

CCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGC
ACACACTGTATTGGACCGGCCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAG
CTGAATGGCCAGGCGGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCAC
ACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCC
CGACACCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGT
CTGATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTCTCACGAGAT
CATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAA
CTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGG
AGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATC
ACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGC
AGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCA
GGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTC
ATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGTGGAGAACC
TGAATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCA
GTTCCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGTGAACGACTATCCAGCAG
AGAAAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCC
AAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCATATACATCTAAGATC
GATCCCTGACCGGCTTCGTGGACCCCTTCGTGTGAAAACCATCAAGAATCACGAGAG
CCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACT
TCATCCTGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTA
TGCTGCGATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
CCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTACCGGCGAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
GTGTTCCAGGGATGGCTCCAACATCCTGCCAAGCTGCTGGAGAATGACGATTCTCACGC
CATCGACACGATGGTGGCCCTGATCCGACGCTGTGCGAGATGCGGAACCTCCAATGCC
GCCACAGCGGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCG
ACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACGCGGATGCCAATGGCGCCTACCA
CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACTGAAGGAGAGCAAGGATCTGAAG
CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaa
aaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaagGGAGCGGCCGCACTCGAGCACC
ACCACCACCACCACTGA SEQ ID NO:428)

AAS1935: pET-28b-heAsCas12a-HF1(E174R/N282A/S542R/K548R)-NLS-6xHis

Codones humanos optimizados de Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para eAsCas12a- HF1 (E174R/N282A/S542R/K548R) en minúsculas y doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS doble subrayado, glicina insertada subrayado punteado, NLS de nucleoplasmina (aaaaggccggcgccac- gaaaaaggccccaggcaaaaaagaaaaag, SEQ ID NO: 21) en minúsculas, secuencias enlazadoras en *cursiva*, 6xHis en **negrita**

5

ATGGGGACACAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTG
 CGGTTTGAGCTGATCCACAGGGCAAGACCTGAAGCACATCCAGGAGCAGGGCTTCA
 TCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCG
 GATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAAC
 CTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCC
 TGATCGAGGAGCAGGCCACATATCGCAATGCCATCCAGACTACTTCATCGGCCGGACA
 GACAACTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAA
 GGCCGAGCTGTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCAACCGAG
 CACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTAT
agaAACAGGAAGAACGTGTTACGCGCCGAGGATATCAGCACAGCCATCCACACCGCAT
 CGTGCAGGACAACCTCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCTGATCAC
 CGCCGTGCCAGCCTGCGGGAGCACTTTGAGAAGCTGAAGAAGGCCATCGGCATCTTC
 GTGAGCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCAAGCTGCTGACACAG
 ACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCG
 AGAAGATCAAGGGCCTGAACGAGGTGCTGggcCTGGCCATCCAGAAGAATGATGAGACA
 GCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTC
 CGATAGGAACACCCTGTCTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCC
 AGTCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAAGCTGCTGGAGACAGCCGAG
 GCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAA
 GCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTG
 TATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAGTCTGCCAAGGAGAAGG
 TGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAAGG
 CAAGGAGCTGAGCGAGGCCCTTCAAGCAGAAAACCGAGATCCTGTCCACGCACAC
 GCCGCCCTGGATCAGCCACTGCCTACAACCCCTGAAGAAGCAGGAGGAGAAGGAGATCC
 TGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGT
 GGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTG
 GAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCC
 CTACTCCGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCaggGGCTGGG
 ACGTGAATaggGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAAGGCCTGTACTATC
 TGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGA
 GAAAACCAAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCAAGA
 TGATCCCAAAGTGACGACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACA
 ACCCCCATCCTGCTGTCCAACAATTTATCGAGCCTCTGGAGATCACAAGGAGATCTAC
 GACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAAC
 CGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGAT
 TTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCT
 CAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAG
 CTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTAC
 CTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCA
 CACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGC

TGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACA
 CCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCC
 GACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTC
 TGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATC
 ATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCCTATCACACTGAAC
 TATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGA
 GCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCA
 CAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCA
 GTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAGGAGAGGGTGGCAGCAAGGCAG
 GCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCA
 TCCACGAGATCGTGACCTGATGATCCACTACCAGGCCGTGGTGGTCTGGAGAACCT
 GAATTTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAG
 TTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTCTGAAGGACTATCCAGCAGA
 GAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCA
 AGATGGGCACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCCATATACATCTAAGATCG
 ATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGC
 CGCAAGCACTTCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTT
 CATCCTGCACCTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTAT
 GCCTGCATGGGATATCGTGTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
 CCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTACCCGGCAG
 ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
 GTGTTCAAGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGC
 CATCGACACCGATGGTGGCCCTGATCCGCAGCGTGTGTCAGATGCCGAACTCCAATGCC
 GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCG
 ACTCCCGGTTTTAGAACCCAGAGTGGCCAAATGGACGCCGATGCCAATGGCGCTACCA
 CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
 CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaa
 aaggccggccggccagcaaaaaggccggccaggcaaaaagaaaaagGGAGCGCGCGCACTCGAGCACC
 ACCACCACCACCCTGA SEQ ID NO: 429)

Secuencia de nucleótidos de SQT1665 pCAG-humanLbCpf1-NLS-3xHA

Codón humano optimizado LbCpf1 en letra normal, nts 1-3684), NLS (aaaaggccggccggccacg-
 aaaaaggccggccaggcaaaaagaaaaag, SEQ ID NO: 21) en minúsculas, marcador 3xHA
 5 (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTC-
 CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**, *secuencia enlazadora en cursiva*

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCTGAGGT
 TCAAGGCCATCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGGT
 GAGGACGAGAAGAGAGCCGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTACTA
 TCTGTCTTTTATCAACGACGTGCTGCACAGCATCAAGCTGAAGAATCTGAACAATTACATC

AGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGGAGAATAAGGAGCTGGAGAACCTGGA
 GATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACGAGGGCTACAAGTCCC
 TGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCTCTGGACGATAAGGACGAGA
 TCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCTTTGATAACAG
 AGAGAATATGTTTTCCGAGGAGGGCCAAGAGCACATCCATCGCCTTCAGGTGTATCAACGA
 GAATCTGACCCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGAT
 AAGCACGAGGTGCAGGAGATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGA
 TTTCTTTGAGGGCGAGTTCTTTAACTTTGTGCTGACACAGGAGGGCATCGACGTGTATAAC
 GCCATCATCGCGGGCTTCGTGACCGAGAGCGCGAGAAGATCAAGGGCCTGAACGAGTA
 CATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAGTTTAAGCCACTGTATAAG
 CAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAGGGCTATACATCCGATGA
 GGAGGTGCTGGAGGTGTTTGAACACCCCTGAACAAGAACAGCGAGATCTTCAGCTCCAT
 CAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGACGAGTACTCTAGCGCCGGCATCTTTGT
 GAAGAACGGCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCGAGTGGAACGTGAT
 CCGGGACAAGTGGAATGCCGAGTATGACGATATCCACCTGAAGAAGAAGGCCGTGGTGA
 CCGAGAAGTACGAGGACGATCGGAGAAAGTCCTTCAAGAAGATCGGCTCCTTTTCTCTGG
 AGCAGCTGCAGGAGTACGCCGACGCCGATCTGTCTGTGGTGGAGAAGCTGAAGGAGATC
 ATCATCCAGAAGGTGGATGAGATCTACAAGGTGTATGGCTCCTCTGAGAAGCTGTTTCGAC
 CCCGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAACGACGCCGTGGTGGCCATCATGAA
 GGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTACATCAAGGCCCTCTTTGGCGAGGG
 CAAGGAGACAAACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTGGCCTACGACATCCT
 GCTGAAGGTGGACCACATCTACGATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTC
 TAAGGATAAGTTCAAGCTGTATTTTCAGAACCCCTCAGTTCATGGCGGGCTGGGACAAGGA
 TAAGGAGACAGACTATCGGGCCACCATCCTGAGATACGGCTCCAAGTACTATCTGGCCAT
 CATGGATAAGAAGTACGCCAAGTGCCCTGCAGAAGATCGACAAGGACGATGTGAACGGCAA
 TTACGAGAAGATCAACTATAAGCTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTT
 CTTTTCTAAGAAGTGATGGCCTACTATAACCCCGAGCGAGGACATCCAGAAGATCTACAAG
 AATGGCACATTCAAGAAGGGCGATATGTTAACCTGAATGACTGTCACAAGCTGATCGACT
 TCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAATGCCCTACGATTTCAACTTTTCT
 GAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGAGGAGCAGGGCTA
 TAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGGCA
 AGCTGTATATGTTCCAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAA
 TCTGCACACCATGTACTTCAAGCTGCTGTTTGACGAGAACAATCACGGACAGATCAGGCT
 GAGCGGAGGAGCAGAGCTGTTTCATGAGGCGCGCCTCCCTGAAGAAGGAGGAGCTGGTG
 GTGACCCAGCCAACTCCCTATCGCCAACAAGAATCCAGATAATCCCAAGAAAACCACA
 ACCCTGTCCTACGACGTGTATAAGGATAAGAGGTTTTCTGAGGACCAGTACGAGCTGCAC
 ATCCCAATCGCCATCAATAAGTGCCCAAGAACATCTTCAAGATCAATACAGAGGTGCGC
 GTGCTGCTGAAGCACGACGATAACCCCTATGTGATCGGCATCGATAGGGGCGAGCGCAA
 TCTGCTGTATATCGTGGTGGTGGACGGCAAGGGCAACATCGTGGAGCAGTATTCCTGAA
 CGAGATCATCAACAACCTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTGGAC

AAGAAGGAGAAGGAGAGGTTGAGGGCCCGCCAGAACTGGACCTCCATCGAGAATATCAA
GGAGCTGAAGGCCGGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGA
AGTACGATGCCGTGATCGCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCGCGTGA
AGGTGGAGAAGCAGGTGTATCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACTACA
TGGTGGACAAGAAGTCTAATCCTTGTGCAACAGCGCGCCCTGAAGGGCTATCAGATCA
CCAATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATCTTTTACATCCC
TGCCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCTGAAAACCAA
GTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGATCATGTACGTG
CCCGAGGAGGATCTGTTGAGTTTGCCCTGGACTATAAGAACTTCTCTCGCACAGACGCC
GATTACATCAAGAAGTGAAGCTGTACTCCTACGGCAACCGGATCAGAATCTTCCGGAAT
CCTAAGAAGAACAACGTGTTGACTGGGAGGAGGTGTGCTGACCCAGCCCTATAAGGA
GCTGTTCAACAAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGA
GCAGTCCGACAAAGGCCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGCTGCAGAT
GCGGAACAGCATCACAGGCCGACCCGACGTGGATTTTCTGATCAGCCCTGTGAAGAACTC
CGACGGCATCTTCTACGATAGCCGGAATATGAGGCCAGGAGAATGCCATCCTGCCAAA
GAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCTGTGGGCCATCGGCC
AGTTCAAGAAGGCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCTAACAAGG
AGTGGCTGGAGTACGCCAGACGAGCTGAAGCAGCAaaaggccggcgccagaaaaaggccggcca
ggcaaaaaagaaaaagGGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGT
GCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID NO:10)

Secuencia de aminoácidos de LbCpf1-NLS-3xHA

LbCpf1 en texto normal (AAs 1-1228), NLS (krpaatkkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA (YPY-DVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRYYL
SFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKONEGYKSLFKKDHIETI
LPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIARFCINENLTRYISNMDFEK
VDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTEGIDVYNAIIGGFVTESEGEKIKGLNEYI
NLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRTNLNKNSEIFSSIKKLEKL
FKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKAVVTEKYEDDRRKSF
KKIGSFSLEQLQEYADADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDVAVAI
MKDLLDSVKSFENYKAFEGGKETNRDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDK
FKLYFQNPQFMGGWQDKKETDYRATILRYGSKYYLAIMDKKYAKCLOKIDKDDVNGNYEKINY
KLLPGPNKMLPKVFFSKKWMAYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYP
KWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFS
DKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFMRRASLKKEELVHPANSPIANKNPDPNP
KKTTLTSDYVYKDKRFSEDQYELHIPIANKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLL
YIVVYDGKGNIVEQYSLNEIINNFGIRIKTDYHSLDDKKEKERFEARQNWTSIENIKELKAGYIS
QVVHKICELVEKYDAVIALEDLNSGFKNRVRKVEKQVYQKFEKMLIDKLNVMVDKKSNPCATG
GALKGYQITNKFSFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRI

MYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKEL
FNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMQMRNSITGRDVFISPVKNSDGIFY
DSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKADEKLDKVKIAISNKEWLEYAQTSTV
KHkrpaatkkagqakkkkGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:11)

Secuencia de nucleótidos de AAS1472 pCAG-humanFnCpf1-NLS-3xHA

Codón humano optimizado FnCpf1 en letra normal, nts 1-3900), NLS (aaaaggccggcgccacg-
 aaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, marcador 3xHA
 (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGGCATACCCATATGATGTCCCCGA
 CGCC, SEQ ID NO:5) en **NEGRITA**

5

ATGAGCATCTACCAGGAGTTCTGTCACAAAGTATTCACTGAGTAAGACACTGCGGT
 TCGAGCTGATCCACAGGGCAAGACACTGGAGAACATCAAGGCCCGAGGCGCTGATTCTG
 GACGATGAGAAGCGGGCAAAAGACTATAAGAAAGCCAAGCAGATCATTGATAATACCAC
 CAGTTCTTTATCGAGGAAATTCTGAGCTCCGTGTGCATCAGTGAGGATCTGCTGCAGAATT
 ACTCAGACGTGTACTTCAAGCTGAAGAAGAGCGACGATGACAACCTGCAGAAAGGACTTCA
 AGTCCGCCAAGGACACCATCAAGAAACAGATTAGCGAGTACATCAAGGACTCCGAAAAGT
 TTAATAATCTGTTCAACCAGAATCTGATCGATGCTAAGAAAGGCCAGGAGTCCGACCTGAT
 CCTGTGGCTGAAACAGTCTAAGGACAATGGGATTGAAGTGTCAAGGCTAACTCCGATATC
 ACTGATATTGACGAGGCACTGGAAATCATCAAGAGCTTCAAGGGATGGACCACATACCTTA
 AAGGCTTCCACGABAACCGCAAGAAGCTGTACTCCAGCAACGACATTCTACCTCCATCA
 TCTACCGAATCGTCGATGACAATCTGCCAAAGTTCTGGAGAAACAAGGCCAAATATGAATC
 TCTGAAGGACAAAGCTCCCGAGGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGA
 GGAAGTACATTCGATATCGACTATAAGACTAGCGAGGTGAACCCAGAGGGTCTTTTCCCT
 GGACGAGGTGTTTGAAATCGCCAAATTTCAACAATTACCTGAACAGTCCGGCATTACTAAA
 TTCAATACCATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAAGCGCAAGGGAATT
 AACGAATACATCAATCTGTATAGCCAGCAGATCAACGACAAAACCTCTGAAGAAATACAAGA
 TGTCTGTGCTGTTCAAACAGATCCTGAGTGATACCGAGTCCAAGTCTTTTGTGATTGATAA
 ACTGGAAGATGACTCAGACGTGGTCACTACCATGCAGAGCTTTTATGAGCAGATCGCCGC
 TTTCAAGACAGTGGAGGAAAAATCTATTAAGGAAACTCTGAGTCTGCTGTTTCGATGACCTG
 AAAGCCCAGAAGCTGGACCTGAGTAAGATCTACTTCAAAAACGATAAGAGTCTGACAGAC
 CTGTACAGCAGGTGTTTGATGACTATTCCGTGATTGGGACCGCCGTCTGGAGTACATT
 ACACAGCAGATCGCTCCAAAGAACCTGGATAATCCCTCTAAGAAAGAGCAGGAACTGATC
 GCTAAGAAAACCGAGAAGGCAAAATATCTGAGTCTGGAAACAATTAAGCTGGCACTGGAG
 GAGTTCAACAAGCACAGGGATATTGACAAACAGTGCCGCTTTGAGGAAATCCTGGCCAAC
 TTCGCAGCCATCCCATGATTTTGTGATGAGATCGCCAGAACAAAGACAATCTGGCTCAGA
 TCAGTATTAAGTACCAGAACCAGGGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATG
 ACGTGAAAGCCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTGCACAAGCTGAAAA
 TCTTCCATATTAGTCAGTCAGAGGATAAGGCTAATATCCTGGATAAAGACGAACACTTCTA
 CCTGGTGTTCGAGGAATGTTACTTCGAGCTGGCAACATTGTCCCCCTGTATAACAAGATT

AGGAACTACATCACACAGAAAGCCTTACTCTGACGAGAAGTTTAACTGAACCTCGAAAATA
GTACCTTGCCCAACGGGTGGGATAAGAACAAGGAGCCTGACAACACAGCTATCCTGTTCA
TCAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAACAATAAGATCTTCGATGA
CAAAGCCATTAAGGAGAACAAGGGGAAGGATACAAGAAAATCGTGATAAGCTGCTGCC
CGGCGCAAATAAGATGCTGCCAAGGTGTTCTTCAGCGCCAAGAGTATCAAATTCACAA
CCATCCGAGGACATCCTGCGGATTAGAAATCACTCAACACATACTAAGAACGGGAGCCCC
CAGAAGGGATATGAGAAATTTGAGTTCAACATCGAGGATTGCAGGAAGTTTATTGACTTCT
ACAAGCAGAGCATCTCCAAACACCCTGAATGGAAGGATTTTGGCTTCCGGTTTTCCGACA
CACAGAGATATAACTCTATCGACGAGTTCTACCGCGAGGTGGAAAATCAGGGGTATAAGC
TGACTTTTGAGAACATTTCTGAAAGTTACATCGACAGCGTGGTCAATCAGGGAAAGCTGTA
CCTGTTCCAGATCTATAACAAAGATTTTTCAGCATACAGCAAGGGCAGACCACAACTGCAT
ACACTGTACTGGAAGGCCCTGTTGATGAGAGGAATCTGCAGGACGTGGTCTATAAACTG
AACGGAGAGGCCGAAGTGTTTTACCGGAAGCAGTCTATTCTTAAGAAAATCACTCACCCA
GCTAAGCAGGCCATCGCTAACAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGAA
TACGATCTGATTAAGGACAAGCGGTTCAACGAAGATAAGTTCTTTTTCATTGTCCATCA
CCATTAACCTCAAGTCAAGCGGCGCTAACAAAGTTCAACGACGAGATCAATCTGCTGCTGAA
GGAAAAAGCAAACGATGTGCACATCCTGAGCATTGACCGAGGAGAGCGGCATCTGCCCT
ACTATACCCTGGTGGATGGCAAAGGGAATATCATTAAAGCAGGATACATTCAACATCATTGG
CAATGACCGGATGAAAACCAACTACCACGATAAACTGGCTGCAATCGAGAAGGATAGAGA
CTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATTAAGGAGATGAAGGAAGGCTATCT
GAGCCAGGTGGTCCATGAGATTGCAAAGCTGGTCATCGAATACAATGCCATTGTGGTGTT
CGAGGATCTGAACCTTCGGCTTTAAGAGGGGGCGCTTTAAGGTGGAAAAACAGGTCTATCA
GAAGCTGGAGAAAATGCTGATCGAAAAGCTGAATTACCTGGTGTTTAAAGATAACGAGTTC
GACAAGACCGGAGGCGTCTGAGAGCCTACCAGCTGACAGCTCCCTTTGAAACTTTCAAG
AAAATGGGAAAACAGACAGGCATCATCTACTATGTGCCAGCCGGATTCACCTCCAAGATCT
GCCCGGTGACCGGCTTTGTCAACAGCTGTACCCCTAAATATGAGTCAGTGAGCAAGTCCC
AGGAATTTTCAGCAAGTTCGATAAGATCTGTTATAATCTGGACAAGGGGTACTTCGAGTT
TTCTTCGATTACAAGAACTTCGGCGACAAGGCCGCTAAGGGGAAATGGACCATTGCCCTC
CTTCGGATCTCGCCTGATCAACTTTGGAATTCGATAAAAACCAATTTGGGACACTAGG
GAGGTGTACCCAAACCAAGGAGCTGGAAAAGCTGCTGAAAGACTACTCTATCGAGTATGGA
CATGGCGAATGCATCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATTTTCGCCAAG
CTGACCTCAGTGCTGAATACAATCCTGCAGATGCCGAACCTCAAAGACCGGGACAGAAGT
GACTATCTGATTAGCCCCGTGGCTGATGTCAACGGAACTTCTTCGACAGCAGACAGGCA
CCCCAAAATATGCCTCAGGATGCAGACGCCAACGGGGCCTACCACATCGGGCTGAAGGG
ACTGATGCTGCTGGGCCGATCAAGAACAATCAGGAGGGGAAGAAGCTGAACCTGGTCA
TTAAGAACGAGGAATACTTCGAGTTTGTCCAGAATAGAAATAACaaaggccggccgccaagaaa
aggccggccaggcaaaaaagaaaaagGGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCT
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID
NO: 17)

Secuencia de aminoácidos de FnCpf1-NLS-3xHA

FnCpf1 en texto normal (AAs 1-1300), NLS (krpaatkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA (YPY-DVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGILDDDEKRAKDYKKAKQIIDKYHQ
 FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKQGESDLILWLKQSKDNGIELFKANSDITDIDEALEIHKSFKGWTTYFKGFHENRKN
 VYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVN
 QRVFSLDEVFEIANFNYYLNQSGITKFNTHGGKFVNGENTKRGINEYINLYSQINDKTLKKY
 KMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKA
 QKLDLSKIYFKNDKSLTDLSSQVFDYVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKT
 AKYLSLETIKLALKEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNGGKK
 DLLQASAEEDVKAIKDLLDGTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLGVMNKKNNKIF
 DDKAIKENKGEYKIVYKLLPGANKMLPKVFFSAKSIKFNPSIEDILRIRNHSTHTKNGSPQK
 GYEKFEFNIEDCRKFIDFYKQSIKHPKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFFEN
 ISESYIDSVVNQKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELEF
 YRKQSIPKKITHPAKEAIAANKNDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
 NDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
 KDRDSARKDWKINNEMKEGYLSQVVEIAKLVEYNAINVVFEDLNFQKRGFRKVEKQVY
 QKLEKMLIEKLNLYLFKDFNEFDKGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
 TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSR
 LINFRNSDKNHNWDREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTI
 LQMRNSKTGTEDYLIISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKN
 NQEGKKLNLVIKNEEYFEVQNRNNkpaatkkaggakkkkGSPYDVPDYAYPYDVPDYAYPYD
 VPDYA (SEQ ID NO:18)

AAS2134: pCAG-hMbCas12a-NLS-3xHA

Codón humano optimizado Moraxella bovoculi 237 Cas12a (MbCas12a) en negro, nucleoplasmina bodocal NLS (aaaaggccgcccagcaaaaaaggccgcccagcaaaaaagaaaaag, SEQ ID NO:21) en minúsculas, secuencias enlazadoras en *cursiva*, marcador 3xHA.

5

(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTC-
 CCCGACTATGCC, SEQ ID NO:5) en **NEGRITA**

ATGCTGTTCCAGGACTTTACCCACCTGTATCCACTGTCCAAGACAGTGAGATTTG
 AGCTGAAGCCCATCGATAGGACCTTGAGCACATCCACGCCAAGAACTTCCTGTCTCAG
 GACGAGACAATGGCCGATATGCACCAGAAAGGTGAAAGTGATCCTGGACGATTACCAACG
 CGACTTCATCGCCGATATGATGGGCGAGGTGAAGCTGACCAAGCTGGCCGAGTTCTATG
 ACGTGACCTGAAGTTTCGGAAGAACCCAAAGGACGATGAGCTGCAGAAGCAGCTGAAG

GATCTGCAGGCCGTGCTGAGAAAGGAGATCGTGAAGCCCATCGGCAATGGCGGCAAGT
 ATAAGGCCGGCTACGACAGGCTGTTCCGGCGCCAAGCTGTTTAAGGACGGCAAGGAGCT
 GGGCGATCTGGCCAAGTTCTGTGATCGCACAGGAGGGAGAGAGCTCCCCAAAGCTGGCC
 CACCTGGCCCACTTCGAGAAGTTTTCCACCTATTTACAGGCTTTACGATAACCGGAAG
 AATATGTATTCTGACGAGGATAAGCACACCGCCATCGCCTACCGCCTGATCCACGAGAA
 CCTGCCCCGGTTTATCGACAATCTGCAGATCCTGACCACAATCAAGCAGAAAGCACTCTG
 CCCTGTACGATCAGATCATCAACGAGCTGACCGCCAGCGGCTGGACGTGTCTCTGGC
 CAGCCACCTGGATGGCTATCACAAGCTGCTGACACAGGAGGGCATCACCGCCTACAATA
 CACTGCTGGGAGGAATCTCCGGAGAGGCAGGCTCTCTAAGATCCAGGGCATCAACGA
 GCTGTATCAATTCTCACCACAACCGAGCTGCCACAAGAGCGAGAGAATCGCCAAGCTGA
 GGCCACTGCACAAGCAGATCCTGTCCGACGGCATGAGCGTGTCTTCTGCCCTCTAAG
 TTTGCCGACGATAGCGAGATGTGCCAGGCCGTGAACGAGTTCTATCGCCACTACGCCGA
 CGTGTTCCGCAAGGTGCAGAGCCTGTTCCGACGGCTTTGACGATCACCAGAAGGATGGC
 ATCTACGTGGAGCACAAGAACCTGAATGAGCTGTCCAAGCAGGCCTTCGGCGACTTTGC
 ACTGCTGGGACGCGTGCTGGACGGATACTATGTGGATGTGGTGAATCCAGAGTTCAACG
 AGCGGTTTGCCAAGGCCAAGACCGACAATGCCAAGGCCAAGCTGACAAAGGAGAAGGA
 TAAGTTCATCAAGGGCGTGCACCTCCCTGGCCTCTCTGGAGCAGGCCATCGAGCACTATA
 CCGCAAGGCCACGACGATGAGAGCGTGCAGGCAGGCAAGCTGGGACAGTACTTCAAGCA
 CGGCCGTGGCCGGAGTGGACAACCCCATCCAGAAGATCCACAACAATCACAGCACCATC
 AAGGGCTTTCTGGAGAGGGAGCGCCCTGCAGGAGAGAGAGCCCTGCCAAAGATCAAGT
 CCGGCAAGAATCCTGAGATGACACAGCTGAGGCAGCTGAAGGAGCTGCTGGATAACGC
 CCTGAATGTGGCCCACTTCGCCAAGCTGCTGACCACAAAGACCACACTGGACAATCAGG
 ATGGCAACTTCTATGGCGAGTTTGGCGTGCTGTACGACGAGCTGGCCAAGATCCCCACC
 CTGTATAACAAGGTGAGAGATTACCTGAGCCAGAAGCCTTTCTCCACCAGAGAAGTACAA
 GCTGAACCTTTGGCAATCCAACACTGCTGAATGGCTGGGACCTGAACAAGGAGAAGGATA
 ATTTCCGGCGTGATCCTGCAGAAGGACGGCTGCTACTATCTGGCCCTGCTGGACAAGGC
 CCACAAGAAGGTGTTTGATAACGCCCTTAATACAGGCAAGAGCATCTATCAGAAGATGAT
 CTATAAGTACCTGGAGGTGAGGAAGCAGTTCCCAAGGTGTTCTTTTCCAAGGAGGCCA
 TCGCCATCAACTACCACCTTTCTAAGGAGCTGGTGGAGATCAAGGACAAGGGCCCGCA
 GAGATCCGACGATGAGCGCCTGAAGCTGTATCGGTTTATCCTGGAGTGTCTGAAGATCC
 ACCCTAAGTACGATAAGAAGTTGAGGGCGCCATCGGCGACATCCAGCTGTTTAAGAAG
 GATAAGAAGGGCAGAGAGGTGCCAATCAGCGAGAAGGACCTGTTGATAAGATCAACG
 GCATCTTTCTAGCAAGCCTAAGCTGGAGATGGAGGACTTCTTTATCGGCGAGTTCAAGA
 GGTATAACCCAAGCCAGGACCTGGTGGATCAGTATAATATCTACAAGAAGATCGACTCC
 AACGATAATCGCAAGAAGGAGAATTTCTACAACAATCACCCCAAGTTTAAGAAGGATCTG
 GTGCGGTACTATTACGAGTCTATGTGCAAGCACGAGGAGTGGGAGGAGAGCTTCGAGTT
 TTCCAAGAAGCTGCAGGACATCGGCTGTTACGTGGATGTGAACGAGCTGTTTACCGAGA
 TCGAGACACGGAGACTGAATTATAAGATCTCCTTCTGCAACATCAATGCCGACTACATCG
 ATGAGCTGGTGGAGCAGGGCCAGCTGTATCTGTTCCAGATCTACAACAAGGACTTTTCC
 CCAAGGCCCCACGGCAAGCCCAATCTGCACACCTGTACTTCAAGGCCCTGTTTTCTGA

GGACAACTGGCCGATCCTATCTATAAGCTGAATGGGAGGGCCAGATCTTCTACAGAA
 AGGCCTCCCTGGACATGAACGAGACAACAATCCACAGGGCCGCCAGGTGCTGGAGAA
 CAAGAATCCCGATAATCCTAAGAAGAGACAGTTCTGTACGACATCATCAAGGATAAGAG
 GTACACACAGGACAAGTTTCATGCTGCACGTGCCAATCAACATGAACCTTTGGCGTGACGG
 GCATGACAATCAAGGAGTTCAATAAGAAGGTGAACCAAGTCTATCCAGCAGTATGACGAG
 GTGAACGTGATCGGCATCGATCGGGGCGAGAGACACCTGCTGTACCTGACCGTGATCA
 ATAGCAAGGGCGAGATCCTGGAGCAGTGTTCCTGAACGACATCACCACAGCCTCTGCC
 AATGGCACACAGATGACCACACCTTACCACAAGATCCTGGATAAGAGGGAGATCGAGCG
 CCTGAACGCCCGGGTGGGATGGGGCGAGATCGAGACAATCAAGGAGCTGAAGTCTGGC
 TATCTGAGCCACGTGGTGCACCAGATCAGCCAGCTGATGCTGAAGTACAACGCCATCGT
 GGTGCTGGAGGACCTGAATTTTCGGCTTTAAGAGGGGCCGCTTTAAGGTGGAGAAGCAG
 ATCTATCAGAACTTCGAGAATGCCCTGATCAAGAAGCTGAACCACTGGTGTGAAGGA
 CAAGGCCGACGATGAGATCGGCTCTTACAAGAATGCCCTGCAGCTGACCAACAATTTCA
 CAGATCTGAAGAGCATCGGCAAGCAGACCGGCTTCTGTTTTATGTGCCCCGCTGGAAC
 ACCTCTAAGATCGACCCTGAGACAGGCTTTGTGGATCTGCTGAAGCCAAGATACGAGAA
 CATCGCCCAGAGCCAGGCCCTTCTTTGGCAAGTTCGACAAGATCTGCTATAATGCCGACA
 AGGATTACTTCGAGTTTCACATCGACTACGCCAAGTTTACCGATAAGGCCAAGAATAGCC
 GCCAGATCTGGACAATCTGTTCCACGGCGACAAGCGGTACGTGTACGATAAGACAGCC
 AACCAGAATAAGGGCGCCGCCAAGGGCATCAACGTGAATGATGAGCTGAAGTCCCTGTT
 CGCCCCGCCACCACATCAACGAGAAGCAGCCCAACCTGGTCATGGACATCTGCCAGAAC
 AATGATAAGGAGTTTCACAAGTCTCTGATGTACCTGCTGAAAACCTGCTGGCCCTGCG
 GTACAGCAACGCCCTCTCTGACGAGGATTTATCCTGTCCCCCGTGGCAAACGACGAGG
 GCGTGTCTTTAATAGCGCCCTGGCCGACGATACACAGCCTCAGAATGCCGATGCCAAC
 GGCGCCTACACATCGCCCCGAAGGGGCTGTGGCTGCTGAATGAGCTGAAGAACTCCG
 ACGATCTGAACAAGGTGAAGCTGGCCATCGACAATCAGACCTGGCTGAATTTGCCCCAG
 AACAGGaaaaggccggccacgaaaaaggccggccaggcaaaaagaaaaagGGATCCT**ACCCATA**
CGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGAT
GTCCCCGACTATGCCTAA (SEQ ID NO: 430)

Secuencia de nucleótidos de (RTW876) pCAG-human-dAsCpf1(D908A)-NLS(nucleoplasmina)-3xHA-VPR

Codón humano optimizado dAsCpf1(D908A) en letra normal (NTs 1-3921), Nucleoplasmina NLS en minúsculas (aaaaggccggccggccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO: 21), marcador 3xHA

5 (TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGGTCCGACTATG
 CC, SEQ ID NO:5) en **negrita**, y VPR doble subrayado

ATGACACAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCCG
 TTTGAGCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGA
 GGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCT

ACAAGACCTATGCCGACCACTGCCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAG
 CGCCGCCATCGACTCCTATAGAAAAGGAGAAAACCGAGGAGACAAGGAAGCCCTGATC
 GAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAA
 CCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCG
 AGCTGTTTAAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGA
 GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAA
 CAGGAAGAAGCTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTG
 CAGGACAACCTCCCAAGTTTAAAGGAGAATTGTCACATCTTCACACGCCCTGATCACCGCC
 GTGCCAGCCTGCGGGAGCACTTTGAGAAGCTGAAGAAGGCCATCGGCATCTTCGTGA
 GCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCAGCTGCTGACACAGACCC
 AGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAAGCCACCGAGAA
 GATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCC
 CACATCATCGCTCCCTGCCACACAGATTTCATCCCTGTTTAAGCAGATCCTGTCCGAT
 AGGAACACCCCTGTCTTTCATCCTGGAGGAGTTTAAAGCGCAGCAGGAAGTGATCCAGTC
 CTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCC
 CTGTTTAAAGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCT
 GGAGACAATCAGCAGCGCCCTGTGCCACCACTGGGATACACTGAGGAATGCCCTGTAT
 GAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGC
 AGCGCAGCCTGAAGCAAGGAGATATCAACCTGCAGGAGATCATCTCTGCCCGCAGGCAA
 GGAGCTGAGCGAGGCTTCAAGCAGAAAACCGAGGAGATCCTGTCCCACGCACACGCC
 GCCCTGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
 AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGA
 TGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAG
 ATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTA
 CTCCGTGGAGAAGTTCAAGCTGAACCTTTCAGATGCCTACACTGGCCTCTGGCTGGGAGC
 TGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAAGGCCCTGTACTATCTG
 GGCATCATGCCAAAGCAGAAGGGCCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGA
 AAACCGAGCAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATG
 ATCCCAAAGTGACGCAOCCAGCTGAAGGCCGTGACAGCCCACTTTTCAGACCCACACAAC
 CCCCATCCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACG
 ACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACC
 GGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATT
 TTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTC
 AGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAGC
 TTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACC
 TGTTCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCAC
 ACACTGTATTGGACCGCCTGTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCT
 GAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACAC
 CGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCG
 ACACCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCT

GATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTCTCACGAGATCA
TCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAAGT
ATCAGGCGGCCAATTCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAG
CACCCCGAGACAOCTATCATCGGCATCGCCCGGGGCGAGAGAACTGATCTATATCAC
AGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAG
TTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGG
CCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATC
CACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTCTGGAGAACCTGA
ATTTCCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCCTGTACCAGCAGTT
CGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGA
AAGTGGGAGGCGTCTGAACCCATACCAGCTGACAGACCAGTTCACTCTCTTTGCCAAG
ATGGGCGACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCATATACATCTAAGATCGAT
CCCTGACCGGCTTCGTGGACCCCTTCGTGTGAAAACCATCAAGAATCACGAGAGCC
GCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTC
ATCCTGCACCTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTAT
GCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
CCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTACCCGGCAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
GTGTTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCAGGC
CATCGACACCATGGTGGCCCTGATCCGCAGCGTGTCTGCAGATGCGGAACCTCAATGCC
GCCACAGGGCAGGACTATATCAACAGCCCGTCCGCGATCTGAATGGCGTGTCTTCG
ACTCCCGGTTTCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCA
CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
CTCGAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCCCAACaa
aaggcgggogggcaogaaaaaggcggccaggcaaaaaagaaaaagGGATCCT**ACCCATACGATGTTCC**
AGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACT
ATGCCGGAAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCT
GGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCCGGATG
CCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTGACCTG
GACATGCTGATTAAGTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGC
CAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACC
IACGAGACATTCAGAGCATCATGAAGAAGTCCCCCTTCAGCGCCCCACCGACCCTAG
ACCTCCACCTAGAAGAAATCGCCGTGCCAGCAGATCCAGCGCCAGCGTGCCAAAACCT
GCCCGCCAGCCTTACCCTTCAACAGCAGCCTGAGCACCATCAACTACGACGAGTTCCC
TACCATGGTGTTCGCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCTC
CTCAGGTGCTCCCTCAGGCTCCTGCTCTCCACCGCTCCAGCCTATGGTGTCTGCATG
GCTCAGGCACCGACCCGCTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTC
CACCAGCCCCATAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCA
GCTGCAGTTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCTGCC
GTGTTACCGACCTGGCCAGCGTGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGG

GCATCCCTGTGGCCCTCACACCACCGAGCCCATGCTGATGGAATACCCCGAGGCCAT
CACCCGGCTCGTGACAGGCGCTCAGAGGCTCCTGATCCAGCTCCTGCCCCCTCTGGGA
GCACCAGGCGCTGCCTAATGGAAGTCTGCTGCGGACGAGGACTTCAGCTCTATCGCCG
ATATGGATTTCTCAGCCTTGCTGGGCTCTGCGAGCGGCAGCCGGGATTCCAGGGAAGG
GATGTTTTTGCGAAGCCTGAGGCGCGCTCCGCTATTAGTGACGTGTTTGAGGGCCGC
GAGGTGTGCCAGCCAAAACGAATCCGGCCATTTTCATCCTCCAGGAAGTCCATGGGCCAA
CCGCCCCTCCCGCCAGCCTGCGACCAACACCAACCGGTCCAGTACATGAGCCAGTC
GGGTCAGTGACCCCGGCACAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGACTC
CCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGC
CTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGC
CAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACACACT
TGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGA
TTCTGGATACCTTCTGAACGACGAGTGCCCTCTTGATGCCATGCATATCAGCACAGGA
CTGTCCATCTTCGACACATCTCTGTTTAA SEQ ID NO: 431)

Secuencia de aminoácidos de dAsCpf1(0908A)-NLS(nucleoplasmina)-3xHA-VPR

AsCpf1 en letra normal (AAs 1-1306), NLS(nucleoplasmina) (krpaatkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA (YPYDVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**, y VPR doble subrayado

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNHDHYKELKPIIDRIYK
TYADQCLQLVQLDWEENLSAIDSYRKEKTEETRNALIEEQATYRNAIHQYFIGRTDNLTDANK
RHAIEYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSADIS
TAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQ
TGIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKILSDRNTLSFI
LEEFKSDEEVIQSFCKYKTLRNENVLETAELFNEINLSDLTIFISHKKLETISSALCDHWDT
LRNALYERRISELTGKITSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAA
LDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVDESNEVDPEFSARLTGKILEMEPSLS
FYNKARNYATKKPYSVEKFLNFQMPTLASGWDVNKEKNNGAILFVNGLYYLGIMPKQKG
RYKALSFEPEKTSEGFDMYYDYFPDAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLE
ITKEIYDLNPEKEPKKQATAYAKKTGDQKGYREALCKWDFTRDFLSKYTKTTSIDLSSLRPS
SOYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHGKPNLHLY
WTGLFSPENLAKTSIKLNGQAEFYRPKSRMKRMAHRLGEKMLNKKLDQKTPIDTLYQEL
YDYVNHRLSHDLSDEARALLPNVITKEVSHEIKDRRFTSDKFFFHVPITLNYQAANSKPFNQ
RVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSNLTQQFDYQKKLDNREKERVAA
QAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNFGFKSKRTGIAEKAVYQQFEKM
LIDKLNCLVLKDYPAEKVGGVLNPHYQLTDQFTSFAKMGTSQGFIFYVPAPYTSKIDFLTGFVD
PFVWTKIKNHESRKHFLEGDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKN
ETQFDAKGTPIAGKRIVPVIEHNRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLEND
SHAITMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWFPMADANGAYH

IALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNkrapatkkagqakkkkGSYPYDVPDYAYP
YDVPDYAYPYDVPDYAGSEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDL
DMLGSDALDDFDLDMLINSRSSGSFKKRKVKGSQYLPDTPDRHRIEEKRKRTYETFKSIMKK
SPFSGPTDPRPPRRIAVPSRSSASVPKPAPQPYPTSSSLSTINYDEFPTMVFPSSGQISQASA
LAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPLAPGPPQAVAPPAPKPTQAGEGTLSEA
LLQLQFDDDELGALLGNSTDAVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAIR
LVTGAQRPPDPAPAPLGAPELNGLLSGDEDESSIAMDFEALLGSGSGSRDSREGMELPK
PEAGSAISDVFEGREVCQPKRIRPFHPGSPWANRPLASLAPTPTGPVHEPVGSLTPAPVP
QPLDPAPAVTPEASHLLEDPEETSQAVKALREMAOTVPOKEEAICGOMDLSHPPPRGHL
DELTTTLESMTEDLNLDSPLTPELNEILDTEFLNDECLLHAMHISTGLSIFDTSLE_SEQ ID NO:
432)

Secuencia de nucleótidos de (RTW776) pCAG-humana-dAsCpf1(0908A)triplevariante(E174R/S542R/K548R)-
NLS(núcleo- plasmína)-3xHA-VPR

- 5 Codón humano optimizado dAsCpf1(D908A) en letra normal (NTs 1-3921), Nucleoplasmina NLS en minúsculas
(aaaaggccgcccgcacgaaaaaggccgcccagcgaagaaagaaaag, SEQ ID NO: 21), marcador 3xHA
(TACCCATACGATGTTCCAGATTACGCTTACCCCTACGACGTGCCTGATTATGCATACCCAT ATGATGTC-
CCCGACTATGCC, SEQ ID NO:5) en **negrita**, y VPR doble subrayado

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGG
TTTGAGCTGATCCACAGGGCAAGACCCCTGAAGCACATCCAGGAGCAGGGCTTCATCGA
GGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCT
ACAAGACCTATGCCGACCACTGCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAG
CGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATC
GAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGCCGGACAGACAA
CCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCG
AGCTGTTTAAATGGCAAGGTGCTGAAGCAGCTGGGCAACCGTGACCACAACCGAGCACGA
GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAA
CAGGAAGAAGCTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTG
CAGGACAACCTCCCAAGTTTAAAGGAGAATTGTACATCTTCACACGCCCTGATCACCGCC
GTGCCCAGCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGA
GCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTTATAACCGCTGCTGACACAGACCC
AGATCGACCTGTATAACCGAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAA
GATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCC
CACATCATCGCTCCCTGCCACACAGATTATCCCCCTGTTTAAAGCAGATCCTGTCCGAT
AGGAACACCCCTGTCTTTCATCCTGGAGGAGTTTAAAGAGCGACGAGGAAGTGATCCAGTC
CTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCC
CTGTTTAAAGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCT
GGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTAT

GAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGC
 AGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGAGGCAA
 GGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCACGCACACGCC
 GCCCTGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
 AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGA
 TGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAG
 ATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTA
 CTCCGTGGAGAAGTTCAAGCTGAACCTTTCAGATGCCTACACTGGCCAGAGGCTGGGACG
 TGAATAGAGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTG
 GGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGA
 AAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATG
 ATCCCAAAGTGACGACCCAGCTGAAGGCCGTGACAGCCCACTTTTCAGACCCACACAAC
 CCCCCCTCTGTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAGGAGATCTACG
 ACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAAC
 GGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATT
 TTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCCGCCATCCTCTC
 AGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCACATCAGC
 TTCCAGAGAAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACC
 TGTTCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCAC
 ACACTGTATTGGACCGGCTGTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCT
 GAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACAC
 CGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCG
 ACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCT
 GATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCA
 TCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAACT
 ATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAG
 CACCCCGAGACACCTATCATCGGCATCGCCCGGGGCGAGAGAAAACCTGATCTATATCAC
 AGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAG
 TTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGG
 CCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATC
 CACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCCTGA
 ATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTT
 CGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGA
 AAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGACCAGTTACCTCCTTTGCCAAG
 ATGGGCACCCAGTCTGGCTTCCTGTTTACGTGCCTGCCCCATATACATCTAAGATCGAT
 CCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCC
 GCAAGCACTTCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTC
 ATCCTGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTAT
 GCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACC
 CCTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAG

ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATC
 GTGTTCAAGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGC
 CATCGACACCATGGTGGCCCTGATCCGCAGCGTGTGCAGATGCGGAACTCCAATGCC
 GCCACAGGCGAGGACTATATCAACAGCCCCGTGCCCGATCTGAATGGCGTGTGCTTCG
 ACTCCCGGTTTTCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCA
 CATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
 CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaa
 aaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCT**TACCCATACGATGTTCC**
AGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGGACT
ATGCCGGAAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCT
GGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATG
CCCITGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTGACCTG
GACATGCTGATTAACCTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAGTTGGTAGC
CAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACC
TACGAGACATTCAAGAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAG
ACCTCCACCTAGAAGAATCGCCGTGCCCGACGATCCAGCGCCAGCGTGCCAAAACCT
GGCCCCCAGCCTTACCCCTTCAACAGCAGCCTGAGCACCATCAACTACGACGAGTTCC
TACCATGGTGTTCCCGACCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCTC
CTCAGGTGCTGCCCTCAGGCTCCTGCTCTGCACCGCTCCAGCCATGGTGTCTGCAGTG
GCTCAGGCACCGACCCCGTGCCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTC
CACCGCCCTAAACCTACACAGGCCCGCGAGGGCACACTGTCTGAAGCTCTGCTGCA
GCTGCAGTTCCAGCAGGAGGATCTGGGAGCCCTGCTGGCAACAGCACCGATCCTGCC
GTGTTACCGACCTGGCCAGCGTGGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGG
GCATCCCTGTGGCCCTTCACACCACCGAGCCCATGCTGATGGAATACCCCGAGGCCAT
CACCCGGCTCGTGACAGGGGCTCAGAGGCCCTCCTGATCCAGCTCCTGCCCTCTGGGA
GCACAGGCTGCTTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCG
ATATGGATTTCTCAGCCTTCTGCGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGG
GATGTTTTTGCCGAAGCCTGAGGCCGCTCCGCTATTAGTGACGTGTTGAGGGCCGC
GAGGTGTGCCAGCCAAAACGAATCCGGCCATTTTCATCCTCCAGGAAGTCCATGGGCCAA
CCGCCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCAGTACATGAGCCAGTC
GGGTCACTGACCCCGGCACCGAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGACTC
CCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGC
CCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGC
CAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACT
TGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGA
TTCTGGATACCTTCTGAACGACGAGTGCCTCTTGATGCCATGCATATCAGCACAGGA
CTGTCCATCTTCGACACATCTCTGTTT SEQ ID NO: 433)

Secuencia de aminoácidos de dAsCpf1(D908A)triplevariante(E174RIS542RIK548R)-NLS(nucleoplasmina)-3xHA-VPR

- 5 AsCpf1 en letra normal (AAs 1-1307), NLS(nucleoplasmina) (krpaatkkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA (YPYDVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**, y VPR doble subrayado

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNQHYKELKPIIDRIYK
 TYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEOATYRNAIHQDYFIGRTDNLTDAINK
 RHAEIYKGLFKAELFNGKVLKQLGTVTTEHENALLRSFDKFTTYFSGFYRNKKNVSAEDIS
 TAIPHRIVQDNFPKFKENCHIFTRITAVPSLREHFENVKKAIGIFVSTSIIEEVFSFPFYNQLLTQ
 TQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQLSDRNTLSFI
 LEEFKSDEEVIQSFCKYKTLRNENVLETAELFNEINSLDTHIFISHKKLETISSALCDHWDT
 LRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAA
 LDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLS
 FYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNKEKNNGAILFVKNGLYYLGIMPKQKG
 RYKALSFEPTKTESEGFDKMYDYFPAKMPKCSTQLKAVTAHFQTHHTTPILLSNNFIEPLE
 ITKEIYDLNPEKEPKKQATAYAKKTGDQKGYREALCKWDFTRDFLSKYTKTTSIDLSSLRPS
 SQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHGKPNLHTLY
 WTGLFSPENLAKTSIKLNGQAEFYRPKSRMKRMAHRLGEKMLNKKLDQKTPIDTLYGEL
 YDYVNHRLSHDLSDEARALLPNVITKEVSHEIKDRRFTSDKFFFHVPITLNYQAANSPPSKFNQ
 RVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSNTIQQFDYQKKLDNREKERVERAAR
 QAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNFQFKSKRTGIAEKAVYQQFEKM
 LIDKLNCLVLKDYPAEKVGGVLNPNYQLTDQFTSFAKMGTSQSGFLFYVPAPYTSKIDPLTGFDV
 PFVWTKIKNHESRKHFLEGDFDLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKN
 ETQFDAQGTPIAGKRIVPVIEHNRFTGRYRDLYFANELIALLEEKGIVFRDGSNILPKLEND
 SHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCDFSFRQNPWPMDADANGAYH
 IALKGQLLNLHKLKESKDLKLQNGISNQDWLAYIQELRNkrpaalkkagqakkkkGSYPYDVPDYAYP
YDVPDYAYPYDVPDYAGSEASGSGRADALDDFDLMLGSDALDDFDLMLGSDALDDFDL
DMLGSDALDDFDLMLINSRSSGSPKKRKVGSQYLPDTPDRHRIEEKRKRTYETFKSMKK
SPFSGPTDPRPPRRIAVPSRSSASVPKPAPQPYFTSSLSTINYDEFPTMVFPSSGQISQASA
LAPAPPQVLPCAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEA
LLQLQFDDDLGALLGNSTDAVFTDLASVDNSEFQQLNQGIQVAPHTTEPMLMEYFEATR
LVTAQRPPDPAPAPLQAPGLPGLLGGDEDESSADMDFSALLGSGSGSRDSREGMFLPK
PEAGSAISDVFEGREVCQPKRIRPFHPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVP
QPLDPAPAVTPEASHLLIDPDEETSQAVKALREMAITVTPQKEEAAICGQMDLSHPPRRGHL
DELTTTLESMTEDLNLDSPLTPELNEILDITFLNDECLLHAMHISTGLSIFDTSLF SEQ ID
 NO:434)

RTW1017: pCAG-2xNLS-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-3xHA-gs-VPR(deAs-VPR(1.2))
 Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para
 DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúsculas y doble subrayado, codones con
 5 mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS y doble subrayado, glicina insertada
subrayado punteado, nucleoplasmina NLS (aaaaggccgagccgacgaaaaaggccgagccgagcaaaaaagaaaaag, SEQ ID
 NO:21) en minúsculas, secuencias enlazadoras en cursiva, marcador 3xHA
 (YPYDVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**, SV40 NLS en cursiva minúsculas, VP64-
 p65-RTA (VPR) en cursiva y doble subrayado

ATGGGc~~cc~~aaagaaaaagaggaaagtcGGCAGTGGAc~~ct~~aaaaagaacgaaaggttGGGTCAGGTA
 CACAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCCGTTTGAGCTG
 ATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACA
 AGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACC
 TATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCA
 TCGACTCCTATAGAAAGGAGAAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCA
 GGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGACAGACAACCTGACCG
 ATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTT
 AATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCC
 TGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTAT~~agg~~AACAGGAAGA
 ACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCCGATCGTGCAGGACAA
 CTTCCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCCTGATCACCGCCCGTGCCCA
 GCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTC
 CATCGAGGAGGTGTTTTCTTCCTTTTATAACCAGCTGCTGACACAGACCCAGATCGA
 CCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAG
 GGCCTGAACGAGGTGCTGAATCTGCCATCCAGAAGAATGATGAGACAGCCACATCAT
 CGCCTCCCTGCCACACAGATTTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACA
 CCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCTTCTGC
 AAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAA
 CGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAA
 TCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAG
 AATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGC
 CTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGA
 GCGAGGCCCTTCAAGCAGAAAACCGCGAGATCCTGTCCCACGCACACGCCCGCCCTGGA
 TCAGCCACTGCCTACAACCCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAG
 CTGGACAGCCTGCTGGCCCTGTACCACTGCTGGACTGGTTTGCCGTGGATGAGTCCA
 ACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCC
 TTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGA
 GAAGTTCAAGCTGAACCTTCAGATGCCTACACTGGCC~~agg~~GGCTGGGACGTGAAT~~agg~~GA
 GAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGC
 CAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACAGCGA

GGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTG
CAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCATCCTG
CTGTCCAAACAATTTTCATCGAGCCTCTGGAGATCACAAGGAGATCTACGACCTGAACAAT
CCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCCAGCAGA
AGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAG
TATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCCGCCATCCTCTCAGTATAAGGAC
CTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACACATCAGCTTCCAGAGAAT
CGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATC
TATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTG
GACCGGCCTGTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGG
CCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGA
GAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCCTGTACC
AGGAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCTGATGAGGCCAGG
GCCCTGCTGCCAACGTGATCACCAAGGAGGTGTCTCAGGAGATCATCAAGGATAGGC
GCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAAGTATCAGGCCGCCA
ATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACA
CCTATCATCGGCATCggcCGGGGCGAGAGAACTGATCTATATCAGTGATCGACTCC
ACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAA
GAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCCTGGTCTGTGGT
GGGCAACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCTCCACGAGATCGTG
GACCTGATGATCCACTACCAGGCCGTGGTGGTGTGGAGAACCTGAATTTCCGGCTTAA
GAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTGAGAAGATGCTG
ATCGATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCCGT
GCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGT
CTGGCTTCCTGTTTTACGTGCCTGCCCATATACATCTAAGATCGATCCCTGACCGGCT
TCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTG
GAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAA
GATGAACAGAAATCTGCCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGGATA
TCGTGTTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGC
AAGAGAATCGTCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTA
TCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGC
TCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGT
GGCCCTGATCCGCAGCGTGCTGCAGATGCGGAATCCAATGCCGCCACAGGCGAGGAC
TATATCAACAGCCCCGTGCCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAA
CCCAGAGTGGCCAATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGC
CAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTC
CAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggccacgaaaaag
gcggccaggcaaaaagaaaaagGGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCT
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCGGAAGCAGGGC
CAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGT

GACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTGGATGCCCTTGATGACTTTGAC
CTCGACATGCTCGGCAGTGACGCCCTTGATGATTTTGACCTGGACATGCTGATTAAGTC
TAGAAGTTCCGGATCTCCGAAAAAGAAAGCGAAAGTTGGTAGCCAGTACCTGCCCGACA
CCGACGACCGGCACCGGATCGAGGAAAAAGCGGAAGCGGACCTACGAGACATTCAAGAG
CATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCAGCCCTAGACCTCCACCTAGAAGAA
TCGCCGTGCCAGCAGATCCAGCGCCAGCGTGCCAAAACTGCCCGCCAGCCTTACCC
CTTCACCAGCAGCCTGAGCACCATCACTACGACGAGTTCCCTACCATGGTGTCCCA
GCGGCCAGATCTCTCAGGCCCTGCTCTGGCTCCAGCCCCCTCTCAGGTGCTGCCCTCA
GGCTCTGCTCTGCAACAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCACCGACA
CCCGTGCCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCAGCCCCCTAAAC
CTACACAGGCCGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCCAGCA
CGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTACCCAGCCTG
GCCAGCGTGGACAACAGCGAGTTCCAGCAGCTGCTGAACAGGGCATCCCTGTGGCCC
CTCACACCACCGAGCCCATGCTGATGGAATACCCCGAGGCCATACCCGGCTCGTGAC
AGGCGCTCAGAGGCCCTCTGATCCAGCTCTGCCCTCTGGGAGCACCAGGCCCTGCCCT
AATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGC
CTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTGCCGAAG
CCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAA
AACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGGGCCAACCGCCACTCCCCGCC
AGCCTCGCACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGG
CACCACTCCCTCAGCCACTGGATCCAGCGCCCCGCAGTGACTCCCGAGGCCAGTCACCT
GTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCC
GATACTGTGATTTCCCGAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCA
TCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAG
GATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCT
GAACGACGAGTGCCCTCTTGATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACA
CATCTCTGTTTTAA (SEQ ID NO: 435)

RTW1130: pCAG-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-NLS-gs-VPR(deAs-VPR(1.3))Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúsculas y doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS y doble subrayado, secuencias enlazadoras en *cursiva*, SV40 NLS en *minúsculas cursiva*, VP64-p65-RTA (VPR) en cursiva y doble subrayado

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGA
GCTGATCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAG
GACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAA
GACCTATGCCGACCAGTGCCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCC
GCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGG
AGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTG

ACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCT
 GTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
 GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTAT^{gg}AACAGG
 AAGAACGTGTTACGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTGAGG
 ACAACTTCCCCAAGTTTAAGGAGAATTGTACATCTTCACAGGCTGATCACCGCCGTGC
 CCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCAC
 CTCCATCGAGGAGGTGTTTTCTTCCTTTTATAACCAGCTGCTGACACAGACCCAGAT
 CGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATC
 AAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCACAT
 CATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTAAGCAGATCCTGTCCGATAGGA
 ACACCTGTCTTTTCATCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTC
 TGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGT
 TTAACGAGCTGAACGACATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAG
 ACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGC
 GGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCG
 CAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGACGGCAAGGAG
 CTGAGCGAGGCCCTTCAAGCAGAAAACCGAGCAGATCCTGTCCACGCACACGCCGCC
 TGGATCAGCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCT
 CAGCTGGACAGCCTGCTGGGCCTGTACCACTGCTGGAAGTGGTTTGCCGTGGATGAGT
 CCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGA
 GCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAA^{gc}CCTACTCCGT
 GGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCC^{ga}GGCTGGGACGTGAAT^{ag}
^gGAGAAGACAATGGGCCATCCTGTTTGTGAAGAACGGCCTGTAATCTCTGGGCATCA
 TGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAAACCG
 CGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCAAA
 GTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCATC
 CTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAGGAGATCTACGAOCTGAA
 CAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACC
 AGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTACAAGGGATTTTCTGTCC
 AAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAG
 GACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCAGAG
 AATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAG
 ATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTA
 TTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCC
 AGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGG
 AGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCTGT
 ACCAGGAGCTGTACGACTATGTGAATCAGACTGTCCACGACCTGTCTGATGAGGCC
 AGGGCCCTGCTGCCAACGTGATCACCAAGGAGGTGTCTCAGGAGATCATCAAGGATA
 GGCCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAATATCAGGCCG
 CCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCCTACCTGAAGGAGCACCCCGAG

83

GCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATT
TCATCCTCCAGGAAGTCCATGGGGCAACCGCCCACTCCCCGCCAGCCTCGCACCAACA
CCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGC
CACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGA
TGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATCTGTGATTCCC
CAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGG
GCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGAC
TCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCTGAACGACGAGTGCCCT
CTTGCAATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTTTAA
 (SEQ ID NO: 436)

RTW1319: pCAG-2xNLS-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-NLS-gs-VPR(deAs-VPR(1.4))
 Codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para
 DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúsculas y doble subrayado, codones con
 5 mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS y doble subrayado, glicina insertada subrayado
punteado, secuencias enlazadoras en *cursiva*, marcador 3xHA (YPYDVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID
 NO: 8) en **negrita**, SV40 NLS en cursiva minúscula, VP64- p65-RTA (VPR) en cursiva y doble subrayado

ATGGGGCccaaagaaaaagaggaaagtcGGCAGTGGAcctaaaaagaaacgaaaggttGGGTACAGGTACA
 CAGTTCCAGGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGAT
 CCCACAGGGCAAGACCCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAAG
 GCCCCGAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTA
 TGCCGACCACTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATC
 GACTCCTATAGAAAGGAGAAACCGAGGAGACAAAGGAACGCCCTGATCGAGGAGCAGG
 CCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGAT
 GCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTTAA
 TGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCTG
 CTGCGGAGCTTCGACAAGTTTTACAACCTACTTCTCCGGCTTTTATaggAACAGGAAGAAC
 GTGTTCCAGCGCCGAGGATATCAGCACAGCCATCCACACCCGATCGTGACGGACAACCTT
 CCCCAGTTTAAGGAGAATTGTACATCTTCACACGCCTGATCACCGCCGTGCCAGCC
 TGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACTCCATC
 GAGGAGGTGTTTTCTTCTCTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTG
 TATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCC
 TGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCC
 TCCCTGCCACACAGATTATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCCT
 GTCTTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCTTCTGCAAGT
 ACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGA
 GCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCA
 GCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAAT

CTCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTG
 AAGCAAGAGGATATCAACCTGCAGGAGATCATCTCTGCCGAGGCAAGGAGCTGAGCG
 AGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCACGCACACGCCGCCCTGGATCA
 GCCACTGCCTACAACCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTG
 GACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGGCGTGGATGAGTCCAACG
 AGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTC
 TCTCAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAGCCCTACTCCGTGGAGA
 AGTTCAAGCTGAACTTTCAGATGCCTACACTGGCC~~agg~~GGCTGGGACGTGAATAGAGAG
 AAGAACAATGGCGCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCC
 AAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAAACCAGCGAG
 GGCTTTGATAAGATGTAATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGC
 AGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCATCCTGCT
 GTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCC
 TGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAG
 GGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTA
 TACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCT
 GGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCAGAGAATCG
 CCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTAT
 AACAAAGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGAC
 CGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCC
 GAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGCTGGGAGAGA
 AGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCGACACCTGTACCAG
 GAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCTGATGAGGCCAGGGC
 CCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCT
 TTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAATATCAGGCCGCCAATT
 CCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCT
 ATCATCGGCATC~~ggc~~CGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACC
 GGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAA
 GCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCCTGGTCTGTGGTGGG
 CACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTATCCACGAGATCGTGGACC
 TGATGATCCACTACCAGGCCCTGGTGGTGTGGAGAACCTGAATTTGGGCTTTAAGAGC
 AAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCG
 ATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGTCT
 GAACCCATACCAGCTGACAGACCAAGTTACCTCCTTTGCCAAGATGGGCACCCAGTCTG
 GCTTCTGTTTTACGTGCTGCCCCATATACATCTAAGATCGATCCCTGACCGGCTTCG
 TGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAG
 GGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACCTTAAGATG
 AACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGGATATCGT
 GTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTTCATCGCGGCAAG
 AGAATCGTGCCAGTGATCGAGAATCACAGATTACCGGCAGATACCGGGACCTGTATCC

TGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCC
 AACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCAGCCATCGACACGATGGTGGC
 CCTGATCCGCAGCGTCTGCGAGATGCGGAAGTCCAATGCCGCCACAGGCGAGGACTAT
 ATCAACAGCCCCGTGCCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCC
 AGAGTGGCCAATGGACGCCGATGCCAATGGCGCTACACATCGCCCTGAAGGGCCAG
 CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAA
 TCAGGACTGGCTGGCTACATCCAGGAGCTGCCAACGGTGGAAAGCGGAGGGAGTccc
 agaagaagaggaaagtcGGGGGTTCCGGAGGAAGCCAGGCCAGCGGTTCCGGACGGGCTGA
CGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTGACCT
TGACATGCTTGGTTCCGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACG
CCCTTGATGATTTGACCTGGACATGCTGATTAAGTCTAGAAGTTCCGGATCTCCGAAAA
AGAAACGCCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGCCACCGGATCGA
GGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCATCATGAAGAAGTCCCCCTTCA
GCGGCCCCACCGACCTAGACCTCCACCTAGAAGAATCGCCGTGCCAGCAGATCCAG
CGCCAGCGTGCCAAACCTGCCGCCAGCCTTACCCCTTACCAGCAGCCTGAGCACC
ATCAACTACGACGAGTTCCTTACCATGGTGTTCGCCAGCGGCCAGATCTCTCAGGCCTC
TGCTCTGGCTCCAGCCCCCTCTCAGGTGCTGCCCTCAGGCTCTGCTCTGCACCCAGCTC
CAGCCATGGTGTCTGCACTGGCTCAGGCACCGACACCCGTGCCGTGCTGGCTCCTGG
ACCTCCACAGGCTGTGGCTCCACCGCCCTAAACCTACACAGGCCGGCGAGGGCACA
CTGTCTGAAGCTCTGCTGCAGCTGCAGTTTCAGCAGCAGGATCTGGGAGCCCTGCTGG
GAAACAGCACCGATCCTGCCGTGTTACCGACCTGGCCAGCGTGACAAACAGCGAGTT
CCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCCTCACACCACCGAGCCCATGCTG
ATGGAATACCCCCAGGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCCTCCTGATC
CAGCTCCTGCCCTCTGGGAGCACCAGGCCCTGCTTAATGGACTGCTGTCTGGCGACGA
GGACTTCAGCTCTATCGCCGATATGGATTCTCAGCCTTGCTGGGCTCTGGCAGCGGCA
GCCGGGATTCAGGGAAGGGATGTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAG
TGACGTGTTTGGGGCCGCGAGGTGTGCCAGCCAAACGAATCCGGCCATTTCATCCTC
CAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGG
TCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCACTCCCTCAGCCACTGGAT
CCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACTGTTGGAGGATCCCGATGAAGAGA
CGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGA
AGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTG
GATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCT
GACCCCGGAATTGAACGAGATTCTGGATACCTTCTGAACGACGAGTGCCTCTTGCATG
CCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTTTAA (SEQ ID
 NO:437)

- 5 RTW1351: pCAG-rAPOBEC1-gs-XTEN-gs-hdAsCas12a(D908A)-NLS-gs-UGI-NLS(AsBE1.1) APOBEC1 de rata (rAPOBEC1) en **negrita mayúsculas**, glicina insertada subrayado punteado, enlazadores de glicina/serina en *cursiva*, enlazador XTEN en cursiva minúsculas, codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codón modificado para DNasa inactiva (D908A) AsCas12a en minúscula doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en **MAYÚSCULA doble subrayado**, NLS de nucleoplasmina en **negrita minúscula**, UGI en **MAYÚSCULAS NEGRITA CON SUBRAYADO DISCONTINUO**, NLS de SV40 en cursiva y doble subrayado

ATGGGCGAGCTCAGAGACTGGCCCACTGGCTGTGGACCCACATTGAGACGGCGGATC
 GAGCCCCATGAGTTTGAGGTATTCCTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCC
 TGCTTTACGAAATTAATTGGGGGGGCGGCACTCCATTTGGCGACATACATCACAGAA
 CACTAACCAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTTC
 TGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAAT
 GTAGTAGGGCCATCACTGAATTCCGTGTCAAGGTATCCCCACGTCACTCTGTTTATTTAC
 ATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGGCTGCGGGATTG
 ATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACGTCTGGA
 GAAACTTTGTGAATTATAGCCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCT
 GTGGGTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCTGCCTCCTTGTCT
 TCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
 TGTCAATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGTTGAAA
CTGGTG
GTCTTCTGGTGGTTCTAGcggcagcgagactcccggaacctcagagtcctccacacccgaaagtTCGGAG
 GGAGTAGCGGCGGCTCTACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAA
 GACACTGCGGTTTGAGCTGATCCACAGGGCAAGACCTGAAGCACATCCAGGAGCAG
 GGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCAT
 CGATCGGATCTACAAGACCTATGCCGACCAGTGCCCTGCAGCTGGTGCAGCTGGATTGG
 GAGAACCCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAAACCGAGGAGACAAGGA
 ACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGC
 CGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCT
 GTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACA
 ACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGG
 CTTTTATGAGAACAGGAAGAACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCAC
 ACCGCATCGTGCAGGACAACCTTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGC
 CTGATCACCGCCGTGCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCG
 GCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTATAACCAGCTGC
 TGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGC
 AGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAAT
 GATGAGACAGCCACATCATCGCCTCCCTGCCACACAGATTTCATCCCCCTGTTTAAGCA
 GATCCTGTCCGATAGGAACACCTGTCTTTTCATCTGGAGGAGTTTAAGAGCGACGAGG
 AAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAG
 ACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAG
 CCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGG
 AATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAGTCTGCCAA

GGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCT
 GCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCC
 ACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCTGAAAGAACGAGGAGGAGAA
 GGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGG
 TTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCA
 TCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACC
 AAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCTC
 TGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGC
 CTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGA
 GCCACAGAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATG
 CCGCCAAGATGATCCCAAAGTGACGACCCAGCTGAAGGCCGTGACAGCCCACTTTCA
 GACCCACACAACCCCGATCCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACA
 AAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTAC
 GCCAAGAAAACCGGCGACCAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACT
 TCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGC
 GGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTG
 TACCACATCAGCTTCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAG
 GCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACACGGCAAG
 CCTAATCTGCACACACTGTATTGGACCGGCTGTTTTCTCCAGAGAACCTGGCCAAGAC
 AAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGA
 GGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAAC
 CCCAATCCCCGACACCTGTACAGGAGCTGTACGACTATGTGAATCAGAGACTGTCCC
 ACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTC
 TCACGAGATCATCAAGATAGGCGCTTTACCAGCGACAAGTCTTTTTCCACGTGCCTAT
 CACACTGAATATCAGGCCGCCAATTCCTCATCTAAGTTCAACCAGAGGGTGAATGCCT
 ACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCGGGGCGAGAGAAACCTG
 ATCTATATCACAGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACAC
 CATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCA
 GCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGA
 GCCAGGTATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGTCT
 GGAGAACCTGAATTTCCGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTG
 TACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCCTGGTGTGAAGGACTA
 TCCAGCAGAGAAAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGAACAGTTCAACC
 TCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTACGTGCCTGCCCATATACA
 TCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAA
 TCACGAGAGCCGCAAGCACTTCTGGAGGGCTTCGACTTCTGCACTACGACGTGAAAA
 CCGGGCACTTCATCTGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCCCTG
 CCGGCTTTATGCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGC
 CAAGGGCACCCCTTTTCATCGCCGCAAGAGAATCGTCCAGTGATCGAGAATCACAGAT
 TCACCGGCAGATACCGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGA

GAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGAC
 GATTCTCACGCCATCGACACGATGGTGGCCCTGATCCGCAGCGTCTGCAGATGCGGA
 ACTCCAATGCCGCCACAGGCGAGGACTATCAACAGCCCCGTGCGCGATCTGAATGG
 CGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACGCCGATGCCAAT
 GGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCA
 AGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGA
 GCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCTCT
 GGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGATATTATTGAAAAGGAGAC
CGGTAAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCCAGAGGAGGTGGAAGA
AGTCATTGGGAACAAGCCGGAAAGCCGATATACTCGTGCACACCGCCTACGACGAGAG
CACCGACGAGAATGTCATGCTTCTGACTAGCGACGCCCCCTGAATACAAGCCTTGGGCT
CTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAAAGATGCTCTCTGGTGGTTCTC
CCAAGAAGAGAGGAAAGTCTAA (SEQ ID NO:438)

- 5 RTW1295: pCAG-rAPOBEC1-gs-XTEN-gs-hdLbCas12a(D832A)-NLS-gs-UGI-NLS(LbBE1.1) APOBEC1 de rata (rAPOBEC1) en **negrita mayúsculas**, glicina insertada subrayado punteado, enlazadores de glicina/serina en *cursiva*, enlazador XTEN en cursiva minúsculas, codón humano optimizado Lachnospiraceae bacterium ND2006 Cas12a (LbCas12a) en negro, codón modificado para DNasa inactiva (D832A) LbCas12a en minúscula doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS y doble subrayado, NLS de nucleoplasmina en **negrita minúscula**, UGI en MAYÚSCULA NEGRITA CON SUBRAYADO DISCONTINUO, NLS de SV40 en cursiva y doble subrayado

ATGGGCGAGCTCAGAGACTGGCCCACTGGCTGTGGACCCACATTGAGACGGCGGATC
 GAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCC
 TGCTTTACGAAATTAATTGGGGGGCGCGCACTCCATTTGGCGACATACATCACAGAA
 CACTAACAAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTT
 TGTCGGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAAT
 GTAGTAGGGCCATCACTGAATTCCCTGTCAAGGTA TCCCCACGTCACTCTGTTTATTTAC
 ATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGGCCTGCGGGATTGT
 ATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACCTGCTGGA
 GAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCT
 GTGGGTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCTGCCTCCTTGTG
 TCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
 TGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAATCTGGTG
 GTTCTTCTGGTGGTTCTTagcggcagcgagactcccgaggacctcagagtcggccacacccgaaagtTCCGGAG
 GGAGTAGCGGCGGGTCTAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAG
 ACCCTGAGGTTCAAGGCCATCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGC
 GGCTGCTGGTGGAGGACGAGAAGAGAGCCGAGGATTATAAGGGCGTGAAGAAGCTGCT
 GGATCGCTACTATCTGTCTTTTATCAACGACGTGCTGCACAGCATCAAGCTGAAGAATCT

GAACAATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGGAGAATAAGGAGC
 TGGAGAACCTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACGA
 GGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCCTGG
 ACGATAAGGACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACC
 GGCTTCTTTGATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGC
 CTTCAGGTGTATCAACGAGAATCTGACCCGCTACATCTCTAATATGGACATCTTCGAGAA
 GGTGGACGCCATCTTTGATAAGCACGAGGTGCAGGAGATCAAGGAGAAGATCCTGAACA
 GCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTTAACTTTGTGCTGACACAGG
 AGGGCATCGACGTGTATAACGCCATCATCGGCGGCTTCGTGACCGAGAGCGGCGAGAA
 GATCAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGC
 CTAAGTTTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTAC
 GCGAGGGCTATACATCCGATGAGGAGGTGCTGGAGGTGTTAGAAACACCCGAAACA
 AGAACAGCGAGATCTTCAGCTCCATCAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGAC
 GAGTACTCTAGCGCCGGCATCTTTGTGAAGAACGGCCCCGCCATCAGCACAACTCTCCAA
 GGATATCTTCGGCGAGTGGAACGTGATCCGGGACAAGTGGAAATGCCGAGTATGACGAT
 ATCCACCTGAAGAAGAAGGCCGTGGTGACCGAGAAGTACGAGGACGATCGGAGAAAGT
 CCTTCAAGAAGATCGGCTCCTTTTCTCTGGAGCAGCTGCAGGAGTACGCCGACGCCGAT
 CTGTCTGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATCTACAA
 GGTGTATGGCTCCTCTGAGAAGCTGTTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGA
 AGAAGAACGACGCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTC
 GAGAATTACATCAAGGCCTTCTTTGGCGAGGGCAAGGAGACAAACAGGGACGAGTCTTT
 CTATGGCGATTTTGTGCTGGCCTACGACATCCTGCTGAAGGTGGACCACATCTACGATG
 CCATCCGCAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAGCTGTATTTTC
 AGAACCCCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTATCGGGCCAC
 CATCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGT
 GCCTGCAGAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACTATAAG
 CTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTCTAAGAAGTGGATGGC
 CTAATAAATCCAGCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGAAGG
 GCGATATGTTTAACCTGAATGACTGTCAACAAGCTGATCGACTTCTTTAAGGATAGCATCT
 CCCGGTATCCAAAGTGGTCCAATGCCACGATTTCAACTTTTCTGAGACAGAGAAGTATA
 AGGACATCGCCGGCTTTTACAGAGAGGTGGAGGAGCAGGGCTATAAGGTGAGCTTCGA
 GTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGGCAAGCTGTATATGTTT
 CAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCTGCACACCATG
 TACTTCAAGCTGCTGTTTGACGAGAACAATCACGGACAGATCAGGCTGAGCGGAGGAGC
 AGAGCTGTTTCATGAGGCGCCCTCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCC
 AACTCCCTATCGCCAACAAGAATCCAGATAATCCCAAGAAAACCAACCCCTGTCCTAC
 GACGTGTATAAGGATAAGAGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATCGC
 CATCAATAAGTGCCCAAGAACATCTTCAAGATCAATACAGAGGTGCCGCTGCTGCTGA
 AGCACGACGATAACCCCTATGTGATCGGCATCgggAGGGGCGAGCGCAATCTGCTGTAT
 ATCGTGGTGGTGGACGGCAAGGGCAACATCGTGGAGCAGTATTCCTGAACGAGATCA

TCAACAACCTTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGAAG
GAGAAGGAGAGGTTGAGGCCCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAGC
TGAAGGCCGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTAC
GATGCCGTGATCGCCCTGGAGGACCTGAACCTCTGGCTTTAAGAATAGCCGCGTGAAGGT
GGAGAAGCAGGTGTATCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACCTACATGG
TGGACAAGAAGTCTAATCCTTGTGCAACAGGCGGCGCCCTGAAGGGCTATCAGATCACC
AATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATCTTTTACATCCCT
GCCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCTGAAAACCAA
GTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGATCATGTACGT
GCCCCGAGGAGGATCTGTTGAGTTTGCCTGGACTATAAGAACTTCTCTCGCACAGACG
CCGATTACATCAAGAAGTGGAAAGCTGTACTCCTACCGCAACCGGATCAGAATCTTCGG
AATCCTAAGAAGAACAACGTGTTTCTGACTGGGAGGAGGTGTGCTGACCAAGCGCCTATAA
GGAGCTGTTCAACAAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGT
GCGAGCAGTCCGACAAGGCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGCTG
CAGATGCGGAACAGCATCACAGGCGCACCGACGTGGATTTTCTGATCAGCCCTGTGAA
GAACTCCGACGGCATCTTCTACGATAGCCGGAATATGAGGCCAGGAGAATGCCATCC
TGCCAAAGAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCTGTGGGC
CATCGGCCAGTTCAAGAAGGCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATC
TCTAACAAGGAGTGGCTGGAGTACGCCAGACCGAGCGTGAAGCACaaaaggccggcgcca
cgaaaaggccggccaggcaaaaagaaaaagGGATCCTCTGGTGGTTCTGGAGGATCTGGTGGT
TC TACTAATCTGTCAGATATTGAAAAAGGAGACCGGTAAAGCAACTGGTTATCCAGG
AATCCATCTCATGCTCCAGAGAGGCTGGAAGAAGTCATTGGGAACAAGCCGGA
GCGATATACTGTCACACCGCCTACGACGAGAGCACCGACGAGAATGTCATGCTTCT
GACTAGCGACGCCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGG
TGAGAAACAAGATTAAAGATGCTCTCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTCTAA
(SEQ ID NO:439)

RTW1352: pCAG-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-UGI-
NLS(eAsBE1.1)

- 5 APOBEC1 de rata (rAPOBEC1) en **negrita mayúscula**, glicina insertada subrayado punteado, enlazadores de glicina/serina en *cursiva*, enlazador XTEN en minúscula cursiva, codón humano optimizado Acidaminococcus sp. BV3L6Cas12a (AsCas12a) en negro, codones modificados para DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúscula y doble subrayado, codones con mutaciones silenciosas para eliminar los sitios NcoI en MAYÚSCULAS y doble subrayado, NLS de nucleoplasmina en **negrita minúsculas**, UGI en **NEGRITA MAYÚSCULAS CON SUBRAYADO DISCONTINUO**, SV40 NLS~ en *cursiva y doble subrayado*

ATGCGGAGCTCAGAGACTGGCCAGTGGCTGTGGACCCACATTGAGACGGCGGATC
GAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCC
TGCTTTACGAAATTAATTGGGGGGGCCGCACTCCATTTGGCGACATACATCACAGAA

10

CACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTTC
 TGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAAT
 GTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGTCACCTCTGTTTATTTAC
 ATCGCAAGGCTGTACCACCACGCTGACCCCGCAATCGACAAGGCCTGCGGGATTG
 ATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGGA
 GAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCT
 GTGGGTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCGCTCCTTGTG
 TCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
 TGTCAATTACCAGCGACTGCCCCACACATTCTCTGGGCCACCGGGTTGAAAATCTGGTG
 GTTCTTCTGGTGGTTCTAGCGGCGAGAGCTCCGGGACCTAGAGTCCGCCACACCGAAAGTTCGGAG
 GGAGTAGCGGGCGGGTCTACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAA
 GACACTGCGGTTTGAGCTGATCCACAGGGCAAGACCCCTGAAGCAGATCCAGGAGCAG
 GGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCAT
 CGATCGGATCTACAAGACCTATGCCGACCAAGTGCCTGCAGCTGGTGCAGCTGGATTGG
 GAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGA
 ACGCCCTGATCGAGGAGCAGGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGC
 CGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCT
 GTTCAAGGCCGAGCTGTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACA
 ACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGG
 CTTTATAGGAACAGGAAGAAGCTGTTTACGCGCCGAGGATATCAGCACAGCCATCCCACA
 CCGCATCGTGCAGGACAACCTCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCT
 GATCACCGCCGCTGCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGC
 ATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCTTCCTTTTATAACCAGCTGCTG
 ACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAG
 GCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGATGA
 TGAGACAGCCACATCATCGCCTCCCTGCCACACAGATTATCCCCCTGTTTAAGCAGA
 TCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAA
 GTGATCCAGTCTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGAC
 AGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCC
 ACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAA
 TGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAG
 GAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTG
 CCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCGAGGAGATCCTGTCCA
 CGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCCTGAAGAAGCAGGAGGAGAAG
 GAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTT
 TGCCGTGGATGAGTCCAACGAGGTGACCCCGAGTTCTCTGCCCGGCTGACCGGCATC
 AAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAA
 GAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACCTTTCAGATGCCCTACACTGGCCGGG
 GCTGGGACGTGAATAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAAGGCCTG
 TACTATCTGGGCATCATGCCAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCC

CACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTAAGTACTATGACTACTTCCCTGATGCCG
 CCAAGATGATCCCAAAGTGACGACCCAGCTGAAGGCCGTGACAGCCACTTTAGACC
 CACACAACCCCATCCTGCTGTCCAACAATTTATCGAGCCTCTGGAGATCACAAGGA
 GATCTACGACCTGAACAATCCTGAGAAGGAGGCCAAAGAGTTTCAGACAGCCTACGCCA
 AGAAAACCGGCGACCCAGAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCAC
 AAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCCGGC
 ATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACC
 ACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCCAA
 GCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTA
 ATCTGCACACACTGTATTGGACCGGCCCTGTTTCTCCAGAGAACCTGGCCAAGACAAGC
 ATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGAT
 GGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCA
 ATCCCCGACACCTGTACCAGGAGCTGTACGACTATGTGAATCAGAGACTGTCCACGGA
 CCTGTCTGATGAGGCCAGGGCCCTGCTGCCAACGTGATCACCAGGAGGTGTCTCAC
 GAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACA
 CTGAACATCAGGCCGCCAATTCCTCATCTAAGTTCAACCAGAGGGTGAATGCCATCCT
 GAAGGAGCACCCCGAGACACCTATCATCGGCATCGCGGGCGAGAGAAACCTGATCT
 ATATCACAGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATC
 CAGCAGTTTGATTACCAGAAAGCTGGACAACAGGGAGAGAGGGTGGCAGCAA
 GGCAGGCCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCA
 GGTCTACCCAGATCTGGACCTGATGATCCACTACCAGGCCGTGGTGGTCTGGAG
 AACCTGAATTTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCA
 GCAGTTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTCTGAAGGACTATCCAG
 CAGAGAAAGTGGGAGGGCTGCTGAACCCATACCAAGCTGACAGACCAAGTTACCTCCTTT
 GCCAAGATGGGCACCCAGTCTGGCTTCTGTTTACGTGCCTGCCCATATACATCTAA
 GATCGATCCCTGACCGGCTTCTGGACCCCTTCTGTGGAAAACCATCAAGAATCACG
 AGAGCCGCAAGCACTTCTGGAGGGCTTCTGACTTTCTGCACTACGACGTGAAAACCGGC
 GACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGG
 CTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACAGTTTGACGCCAAGG
 GCACCCCTTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCAGAGTTCAAC
 GGCAGATACCGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGG
 GCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAGCTGCTGGAGAATGACGATTCT
 CACGCCATCGACACGATGGTGGCCCTGATCCGACGCTGTGTCAGATGCGGAACCCA
 ATGCCGCCACAGGCCGAGGACTATATCAACAGCCCCGTGCCGATCTGAATGGCGTGTG
 CTTTCGACTCCCGTTTCAGAACCCAGAGTGGCCAATGGACGCCGATGCCAATGGCGCC
 TACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCT
 GAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCCG
 AACaaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCCTGGTGGTTC
 TGGAGGATCTGGTGGTTC**ACTAATCTGTCAGATATTGAAAAAGGAGACCGGTAAG**
CAACTGGTTATCCAGGAATCCATCCTCATGCTCCAGAGGAGGTGGAAGAAGTCATTG

GGAACAAGCCGGAAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACCCGACG
AGAATGTCATGCTTCTGACTAGCGACGCCCTGAATACAAGCCTTGGGCTCTGGTCTAT
ACAGGATAGCAACGGTGAGAACAAGATTAAGATGCTCTCTGGTGGTTCCCCAAGAAG
AAGAGGAAAGTCTAA (SEQ ID NO:440)

RTW1348: pCAG-2xNLS-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-UGI-NLS(eAsBE1.2)

- 5 APOBEC1 de rata (rAPOBEC1) en **negrita mayúscula**, glicina insertada subrayado punteado, SV40 NLS en *cursiva doble subrayado*, enlazadores glicina/serina en *cursiva*, enlazador XTEN en *cursiva minúsculas*, codón humano

optimizado de Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúsculas doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en MAYÚSCULAS doble subrayado, NLS de nucleoplasmina en negrita minúsculas, UGI en NEGRITA MAYÚSCULAS CON SUBRAYADO DISCONTINUO

5

ATGGGCCCAAAGAAAAAGAGGAAAGTCGGCAGTGGACCTAAAAAGAAACGAAGGTTG
GGTCAGGTAGCTCAGAGACTGGCCAGTGGCTGTGGACCCACATTGAGACGGCGGA
TCGAGCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTG
CCTGCTTTACGAAATTAATTGGGGGGGCCGCACTCCATTTGGCGACATACATCACAG
AACACTAACAGCACGTGCAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATT
TCTGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGA
ATGTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGTCACTCTGTTTATT
ACATCGCAAGGCTGTACCACCACGCTGACCCCGCAATCGACAAGGCCTGCGGGATT
TGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTG
GAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCAT
CTGTGGGTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCTGCCTCCTTG
TCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGT
CTTGTCATTACCAGCGACTGCCCCACACATTCTCTGGGCCACCGGGTTGAAATCTGG
TGGTTCTCTGGTGGTTCTAGCGGcagagagactccgggagootcagagtccgocacacccgaaagiTCCGG
AGGGAGTAGCGCGGGTCTACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGC
AAGACACTGCGGTTTGAGCTGATCCACAGGGCAAGACCCCTGAAGCACATCCAGGAGC
AGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCAT
CATCGATCGGATCTACAAGACCTATGCCGACCAAGTGCCCTGCAGCTGGTGCAGCTGGATT
GGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAG
GAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCG
GCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGG
CCTGTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACC
ACAACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTC
CGGCTTTTATAGAAACAGGAAGAACGTGTTTCAGCGCCGAGGATATCAGCACAGCCATCC
CACACCGCATCGTGCAGGACAACCTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACAC

GCCTGATCACCGCCGTGCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCAT
CGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCCTCCCTTTTATAACCAGCT
GCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAG
GCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGA
ATGATGAGACAGCCACATCATCGCCTCCCTGCCACACAGATTATCCCCCTGTTTAAG
CAGATCCTGTCCGATAGGAACACCCCTGTCTTTTCATCCTGGAGGAGTTTAAGAGCGACGA
GGAAGTGATCCAGTCCCTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGG
AGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATC
AGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGA
GGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAGTCTGC
CAAGGAGAAGGTGCAGCGCAGCCTGAAGCAGGAGATATCAACCTGCAGGAGATCATC
TCTGCCCGAGGCAAGGAGCTGAGCGAGGCCCTTCAAGCAGAAAACGAGCGAGATCCTGT
CCCACGCACACGCCGCCCTGGATCAGCCACTGCCACACCTGAAGAAGCAGGAGGA
GAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCCTGTACCACCTGCTGGAC
TGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCG
GCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCC
ACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACCTTCAGATGCCATACACTGGC
CAGGAGGCTGGGACGTGAATAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGG
CCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCG
AGCCACAGAGAAAACGAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGAT
GCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGGCCACTTTC
AGACCCACACAACCCCATCCTGCTGTCCAACAATTTTCATCGAGCCTCTGGAGATCACA
AAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGGCCAAAGAAGTTTCAGACAGCCTA
CGCCAAAGAAAACCGCGCAGCAGAAAGGCTACAGAGAGGCCCTGTGCAAGTGGATCGAG
TTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTG
CGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCTGCT
GTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACA
GGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAA
GCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAAGA
CAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAG
AGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAA
CCCCAATCCCCGACACCCCTGTACCAGGAGCTGTACGACTATGTGAATCAGAGCTGTCC
CACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAGGAGGTGT
CTCAGGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTCCACGTGCCTA
TCACACTGAACATCAGGCCGCCAATTCCTCATTAAGTTCAACCAGAGGGTGAATGCC
TACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCAGCGGGGCGAGAGAAACCT
GATCTATATCACAGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGCCTGAACA
CCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGC
AGCAAGGCAGGCCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTG
AGCCAGGTCTCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGC

TGGAGAACCTGAATTTGGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGT
GTACCAGCAGTTTCGAGAAGATGCTGATCGATAAGCTGAATTCCTGGTGCTGAAGGACT
ATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAAGTTTAC
CTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCTGTTTTACGTGCCTGCCCCATATAC
ATCTAAGATCGATCCCCGTGACCGGCTTCGTGGACCCCTTCGTGTGAAAAACCATCAAGA
ATCACGAGAGCCGCAAGCACTTCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAA
ACCGGCGACTTCATCCTGCACCTTAAGATGAACAGAAATCTGTCTTCAGAGGGGCGCT
GCCCCGGCTTTATGCCTGCATGGGATATCGTGTTTCGAGAAGAACGAGACACAGTTTGACG
CCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGA
TTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGG
AGAAGGGCATCGTGTTCAAGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGAC
GATTCTCACGCCATCGACACGATGGTGGCCCTGATCCGCAGCGTCTGTCAGATGCGGA
ACTCCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGG
CGTGTGCTTCGACTCCCGGTTTCAGAAGCCAGAGTGGCCAAATGGACGCCGATGCCAAT
GGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCA
AGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGA
GCTGCGCAACaaaaggccggcgccatgaaaaaggccggcaggcaaaaagaaaaagGGATCCTCT
GGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGATATTATTGAAAAGGAGAGC
CGGTAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCCAGAGGAGGTGGAAGA
AGTCAATTGGGAACAAGCCGGAAAGCGATATACTCGTGACACCGCCTACGACGAGAG
CACCGACGAGAATGTCATGCTTCTGACTAGCGACGCCCTGAATACAAGCCTTGGGCT
CTGGTCTATACAGGATAGCAACGGTGAGAACAAAGATTAAAGATGCTCTCTGGTGGTTCTC
CCAAGAAGAAGAGGAAAGTCTAA (SEQ ID NO:441)

- 5 RTW1296: pCAG-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-UGI-NLS (eAsBE1.3) APOBEC1 de rata (rAPOBEC1) en **negrita mayúscula**, glicina insertada subrayado punteado, enlazadores de glicina/serina *en cursiva*, enlazador XTEN en *minúscula cursiva*, codón humano optimizado Acidaminococcus sp. BV3L6 Cas12a (AsCas12a) en negro, codones modificados para DNasa inactiva (D908A) eAsCas12a (E174R/S542R/K548R) en minúsculas doble subrayado, codones con mutaciones silenciosas para eliminar sitios NcoI en **negrita subrayado en negro**, UGI en **MAYÚSCULAS NEGRITAS CON SUBRAYADO DISCONTINUO**, SV40 NLS en cursiva doble subrayado

ATGGGCAGCTCAGAGACTGGCCAGTGGCTGTGGACCCACATTGAGACGGCGGATC
GAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCC
TGCTTTACGAAATTAATTGGGGGGGCGGCACTCCATTTGGCGACATACATCACAGAA
CACTAACAAAGCACGTGCAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTTC
TGTCGGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGGCAAT
GTAGTAGGGCCATCACTGAATTCTGTCAAGGTATCCCCACGTCACTCTGTTTATTAC
ATCGCAAGGCTGTACCACCACGCTGACCCCGCAATCGACAAGGCCTGCGGGATTG

ATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGGA
 GAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCATCT
 GTGGGTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCTGCCTCCTTGTG
 TCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
 TGTCATTACCAGCGACTGCCCCACACATTCTCTGGGCCACCGGGTTGAAAATCTGGTG
 GTTCTTCTGGTGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACC
 CGAAAATTCGGAGGGAGTAGCGGCGGGTCTACACAGTTTCGAGGGCTTTACCAACCTG
 TATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCACAGGGCAAGACCTGAAGCA
 CATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAG
 CTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCACTGCCTGCAGCTGGT
 GCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACC
 GAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCAGC
 ACTACTTCATCGGCCGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAG
 ATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAAATGGCAAGGTGCTGAAGCAGCTGGG
 CACCGTGACCACAACCGAGCAGCAGAACGCCCTGCTGCCGAGCTTCGACAAGTTTACA
 ACCTACTTCTCCGGCTTTTATgggAACAGGAAGAACGTGTTTCAGCGCCGAGGATATCAGC
 ACAGCCATCCACACCGCATCGTGCAGGACAACCTCCCAAGTTTAAGGAGAAATTGTCA
 CATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCCGGAGCACTTTGAGAACGTG
 AAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCTTCCCTTTT
 TATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAAT
 CTCTCGGGAGGCAAGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCC
 ATCCAGAAGAATGATGAGACAGCCACATCATCGCCTCCCTGCCACACAGATTCAATCCC
 CCTGTTTAAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAA
 GAGCGACGAGGAAGTGATCCAGTCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGA
 ACGTGCTGGAGACAGCCGAGGCCCTGTTTAAAGAGCTGAACAGCATCGACCTGACACA
 CATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGG
 GATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCCAAGATCAC
 CAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAG
 GAGATCATCTCTGCCGACGCAAGGAGCTGAGCGAGGCCCTCAAGCAGAAAACCAGCG
 AGATCCTGTCCACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAA
 GCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCAC
 CTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCC
 GGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGA
 AATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCT
 AACTGCGCCagaggGCTGGGACGTGAATaggGAGAAGAACAATGGCGCCATCCTGTTTGTG
 AAGAACGGCCTGTACTATCTGGGCATCATGCCAAGCAGAAGGGCAGGTATAAGGCCCT
 GAGCTTCGAGCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACT
 TCCCTGATGCCGCCAAGATGATCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGC
 CCACCTTTCAGACCCACACAACCCCATCCTGCTGTCCAACAATTTATCGAGCCTCTGGA
 GATCACAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGA

CAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTG
 GATCGACTTCACAAGGGATTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTC
 TAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATC
 CCTGTCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTG
 GAGACAGGCCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCA
 CGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCTGTTTTCTCCAGAGAACCTGG
 CCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAG
 GATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGAT
 CAGAAAACCCCAATCCCGACACCTGTACCAGGAGCTGTACGACTATGTGAATCACAG
 ACTGTCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAG
 GAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCAC
 GTGCCTATCACACTGAACATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGT
 GAATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCCGGGCGGAGA
 GAAACCTGATCTATATCACAGTGATCGACTCCACCGCAAGATCCTGGAGCAGCGGAGC
 CTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAG
 GGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGC
 TATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGT
 GGTGCTGGAGAACCCTGAATTTGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAG
 GCCGTGTACCAGCAGTTCCAGAAGATGCTGATCGATAAGCTGAATTGCTGGTCTGAA
 GGACTATCCAGCAGAGAAAGTGGGAGGCGTGTGAACCCATACCAGCTGACAGACCAG
 TTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCTGTTTACGTCGCTGCCCC
 ATATACATCTAAGATCGATCCCTGACCGGCTTCGTGGACCCCTTCGTGTGAAAACCAT
 CAAGAATCAGGAGCGCGCAAGCACTTCTGGAGGGCTTCGACTTTCTGCACTACGACG
 TGAAAACCGGCGACTTCATCCTGCACCTTAAGATGAACAGAAATCTGTCTTCCAGAGGG
 GCTGCGCCGGCTTTATGCTGCAATGGGATATCGTGTTCGAGAAGAAGCAGACACAGTTT
 GACGCCAAGGGCACCCCTTTCATCGCCGCAAGAGAATCGTGCCAGTGATCGAGAATC
 ACAGATTACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTG
 GAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGA
 ATGACGATTCTCACGCCATCGACACCGATGGTGGCCCTGATCCGCAGCGTGTGTCAGAT
 GCGGAACCTCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCGTGCAGGATCTG
 AATGGCGTGTCTTCGACTCCCGGTTTCAGAACCAGAGTGGCCCAATGGACGCCGATG
 CCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGA
 GAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCC
 AGGAGCTGCGCAACTCTGGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGAT
 ATTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCC
 CAGAGGAGGTGGAGAAGTCAATTGGGAACAAGCCGGAAGCGATATACTCGTGACACA
 CCGCCTACGACGAGAGCACCGACGAGAATGTATGCTTCTGACTAGCGACGCCCTG
 AATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAAAGAT
 GCTCTCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTCTAA (SEQ ID NO:442)

Secuencia de nucleótidos de (JG1211) pCAG-humana-dLbCpf1(D832A)-NLS(nucleoplasmina)-3xHA-VPR

Codón humano optimizado dLbCpf1(D832A) en letra normal (NTs 1-3921), Nucleoplasmina NLS en minúsculas (aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO: 21), marcador 3xHA (TACCCATAC GATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTC-CCCGACTATGCC, SEQ ID NO:5) en **negrita**, y VPR doble subrayado

5

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCCTGAGG
 TTCAAGGCCATCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGG
 TGGAGGACGAGAAGAGAGGCCGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTA
 CTATCTGTCTTTTATCAACGACGTGCTGCACAGCATCAAGCTGAAGAACTGAACAATTA
 CATCAGCCTGTTCCGGAAGAAAACCGAACCAGAGAAGGAGAATAAGGAGCTGGAGAAC
 CTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACGAGGGCTACA
 AGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCCCTGGACGATAAG
 GACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCTT
 TGATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGCCTTCAGGT
 GTATCAACGAGAATCTGACCCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGAC
 GCCATCTTTGATAAGCACGAGGTGCAGGAGATCAAGGAGAAGATCCTGAACAGCGACTA
 TGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTTAACTTTGTGCTGACACAGGAGGGCA
 TCGACGTGTATAACGCCATCATCGGCGGCTTCGTGACCGAGAGCGGGCGAGAAGATCAA
 GGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAGT
 TTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAG
 GGCTATACATCCGATGAGGAGGTGCTGGAGGTGTTAGAAACACCCTGAACAAGAACAG
 CGAGATCTTCAGCTCCATCAAGAAGCTGGAGAAGCTGTTCAAGAAATTTGACGAGTACTC
 TAGCGCCGGCATCTTTGTGAAGAACGGCCCCGCCATCAGCACAATCTCCAAGGATATCT
 TCGGCGAGTGGAACGTGATCCGGGACAAGTGGAATGCCGAGTATGACGATATCCACCT
 GAAGAAGAAAGCCGTGGTGACCGAGAAGTACGAGGACGATCGGAGAAAGTCCCTCAAG
 AAGATCGGCTCCCTTTCTCTGGAGCAGCTGCAGGAGTACGCCGACGCCGATCTGTCTGT
 GGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATCTACAAGGTGTATG
 GTCCTCTGAGAAGCTGTTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAAC
 GACGCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTA
 CATCAAGGCCCTTCTTTGGCGAGGGCAAGGAGACAAACAGGGACGAGTCCCTCTATGGC
 GATTTTGTGCTGGCCTACGACATCCTGCTGAAGGTGGACCACATCTACGATGCCATCCG
 CAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAGCTGTATTTTCAGAACCC
 TCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTATCGGGCCACCATCCTG
 AGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGTGCCTGCA
 GAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACTATAAGCTGCTGC
 CCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTCTAAGAAGTGGATGGCCTACTATA
 ACCCCAGCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGAAGGGCGATATG

TTAAACCTGAATGACTGTCAACAGCTGATCGACTTCTTTAAGGATAGCATCTCCCGGTAT
 CCAAAGTGGTCCAATGCCTACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACATC
 GCCGGCTTTTACAGAGAGGTGGAGGAGCAGGGCTATAAGGTGAGCTTCGAGTCTGCCA
 GCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGGCAAGCTGTATATGTTCCAGATCTAT
 AACAAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCTGCACACCATGTACTTCAAG
 CTGCTGTTTGACGAGAACAAATCACGGACAGATCAGGCTGAGCGGAGGAGCAGAGCTGT
 TCATGAGGCGCGCTCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACCTCCCC
 TATCGCCAACAAGAATCCAGATAATCCCAAGAAAACCACAACCCCTGTCTACGACGTGTA
 TAAGGATAAGAGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATCGCCATCAATAA
 GTGCCCAAGAACATCTTCAAGATCAATACAGAGGTGCGCGTGTCTGCTGAAGCACGACG
 ATAACCCCTATGTGATCGGCATCGCCAGGGGCGAGCGCAATCTGCTGTATATCGTGGTG
 GTGGACGGCAAGGGCAACATCGTGGAGCAGTATCCCTGAACGAGATCATCAACAACCTT
 CAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGAAGGAGAAGGAGA
 GGTTCGAGGGCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAGCTGAAGGCCGG
 CTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCGTGA
 TCGCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCCGCTGAAGGTGGAGAAGCA
 GGTGTATCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACTACATGGTGGACAAGA
 AGTCTAATCCTTGTGCAACAGGCGGCGCCCTGAAGGGCTATCAGATCACCAATAAGTTC
 GAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATCTTTTACATCCCTGCCTGGCTG
 ACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCTGAAAACCAAGTATACCAGC
 ATCGCCGATTCCAAGAAGTTTCATCAGCTCCTTTGACAGGATCATGTACGTGCCCGAGGA
 GGATCTGTTTCGAGTTTGCCCTGGACTATAAGAACTTCTCTCGCACAGACGCCGATTACAT
 CAAGAAGTGGAGCTGTACTCCTACGGCAACCGGATCAGAATCTTCCGGAATCCTAAGA
 AGAAACAACGTGTTGACTGGGAGGAGGTGTGCCGTGACCAGCGCCTATAAGGAGCTGTT
 CAACAAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGT
 CCGACAAGGCCCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGGG
 AACAGCATCACAGGCCGCCACCGACGTGGATTTTCTGATCAGCCCTGTGAAGAACTCCGA
 CGGCATCTTCTACGATAGCCGGAACATAGAGGCCAGGAGAATGCCATCCTGCCAAAGA
 ACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCTGTGGGCCATCGGCCA
 GTTCAAGAAGGCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCTAACAAG
 GAGTGGCTGGAGTACGCCCAGACCAGCGTGAAGCACaaagggccggggccagaaaaagggcgg
 ccaggcaaaaaagaaaaagGGATCCT**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGA**
CGTGCTGATTATGCATACCCATATGATGTCCCGACTATGCCGGAAGCGAGGCCAGC
GGTTCGGACGGGCTGACGCAATTGGACGATTTTGATCTGCATATGCTGGGAAGTGACG
CCCTCGATGATTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCG
ACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAA
GTTCCGGATCTCCGAAAAAGAAACGCAAGTTGATAGCCAGTACCTGCCCGACACCGAC
GACCGGCACCGGATCGAGGAAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCATCA
TGAAGAAGTCCCTTCAGCGGCCCCACCGACCTAGACCTCCACCTAGAAGAATCGCC
GTGCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCGCCAGCCTTACCCCTTCA

CCAGCAGCCTGAGCACCATCAACTACGACGAGTTCCTACCATGGTGTTCCTCCAGCGGC
CAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCTCCTCAGGTGCTGCCTCAGGCTCC
TGCTCCTGCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCACCAGCACCCTG
CCTGTGCTGGCTCCTGCACCTCCACAGGCTGTGGCTCCACCAGCCCTAAACCTACACA
GGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACGAGGAT
CTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTACCGACCTGGCCAGCG
TGGACAACAGCGAGTTCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCTCACAC
CACCGAGCCCATGCTGATGGAATACCCGAGGCCATCACCCTGGCTGTGACAGCGCT
CAGAGGCCCTCCTGATCCAGCTCCTGCCCTCTGGGAGCACCAGGCTGCCTAATGGAC
TGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTGCTG
GGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGGCGAAGCCTGAGG
CCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAACGAAT
CCGGCCATTTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTC
GCACCAACACCAACCGGTCCAGTACATGAGCCAGTCCGGTCACTGACCCCGGCCACAG
TCCCTCAGCCACTGGATCCAGCGCCGCGAGTGACTCCCGAGGCCAGTCAGCTGTTGGA
GGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACT
GTGATTTCCCAGAAAGGAAGAGGCTGCAATCTGTGCCAAATCGACCTTTCCCATCCGCC
CCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGA
ACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTGAACGAC
GAGTGCCCTCTTGATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTG
TTTTAA SEQ ID NO:443)

Secuencia de aminoácidos de dLbCpf1(D832A)-NLS(nucleoplasmina)-3xHA-VPR

LbCpf1 en letra normal (AAs 1-1228), NLS(nucleoplasmina) (krpaatkagqakkkk, SEQ ID NO:7) en minúsculas, marcador 3xHA (YPYDVPDYAYPYDVPDYAYPYDVPDYA, SEQ ID NO:8) en **negrita**, y VPR doble subrayado

MSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
 YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDI
 IETILPEFLDDKDEIALVNSFNGFTTFTGFFDNRENMFSEEAKSTSIAPRCINENLTRYISNMDI
 FEKVDAIFDKHEVQEIKEILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESEGEKIKG
 LNEYINLYNQTKQKLPKFKPLYQVLSDRSLSFYGEYTSDEEVLEVFRNTLNKNSEIFSSI
 KKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEVNVIRDKVNAEYDDIHLKKKAVVTEKYE
 DRRRSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDAADFVLEKSLK
 KNDVAVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESYGDFVLAYDILLKVDHIYDAIRNYV
 TQKPYSKDKFLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDD
 VNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKL
 IDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVVEEQGYKVSFESASKKEVDKLVVEEG
 KLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVWH
 PANSPIANKPNPNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD

NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLDKKEKERFEARQN
 WTSIENIKELKAGYISQVWHKICELVEKYDAVIALEDLNSGFKNSRVKEQVYQKFEKMLIDK
 LNYMVDKKSNPCATGGALKGYQITNKFSFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKT
 KYTSIADSKKFISFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKK
 NNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSIT
 GRTOVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKADEK
 LDKVKIAISNKEWLEYAQTSVKHkpaatkkgqakkkkGSYPYDVPDYAYPYDVPDYAYPYDVPD
YAGSEASGSGRADALDDFDLMDLGSDALDDFDLMDLGSDALDDFDLMDLGSDALDDFDL
MLINSRSSGSPKKKRKVGSOYLPDTPDRHRIEEKRKRTYETFKSIMKKSPPSGPTDPRPPPR
RIAVPSRSSASVPKPAQPYPTSSISTINYDEEPTMVFPSSGQISQASALAPAPPQVLPQAPA
PAFAPAMVSALAAQAPAPVPLAPGFPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDELGALL
GNSTDPAVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPA
PLGAPGLPNGLLSGDEDFSSIAMDFESALLGSGSGSRDSREGMFLPKPEAGSAISDVFEGR
EVCQPKRIKPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVQPLDPAPAVTPEAS
HLLEDPEETSQAVKALREMAITVTPQKEEAAICGQMDLSHPPPRGHLELTTLESMTEDL
NLDSPLTPELNEILDTEFLNDECLHAMHISTGLSIEDTSLE SEQ ID NO:444)

Secuencia de nucleótidos de (RTW1008) pCAG-NLS(SV40)x2-rAPOBEC1-gsXTENgslinker-human-dAsCpf1 (D908A)-NLS(nucleoplasmina)-GSlinker-UGI-NLS(SV40)

- 5 Codón humano optimizado dAsCpf1 (D908A) en letra normal (NTs 844-4764), rAPOBEC1 en **negrita** (NTs 67-750), Nucleoplasmina NLS en minúsculas
- (aaaaggccgcccgcacgaaaaaggccgcccagcgaagaaagaaag, SEQ ID NO:21), SV40 NLS en minúsculas (ccaaa-gaaaaagaggaaagtc, cctaaaaagaaacgaaagggt, o cccaagaagaagaggaaagtc, SEQ ID NOs:19, 20, o 22, respectivamente), gsXTENgsl
 10 enlazador(tctggtggttcttctggtggttctagcggcagcgagactcccgggacactcagagtcggccacacccgaaa gttccggaggagtag cggcggg, SEQ ID NO:23) en minúsculas, y UGI doble subrayado

ATGGGCcctaaagaaaaagaggaaagtcGGCAGTGGAcctaaaaagaaacgaaagggtGGGTCAAG
 GTAGCTCAGAGACTGGCCAGTGGCTGTGGACCCACATTGAGACGGCGGATCGAGC
 CCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCT
 TTACGAAATTAATTGGGGGGGCGGCACTCCATTTGGCGACATACATCACAGAACACT
 AACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTC
 CGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAG
 TAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGCTCACTCTGTTTATTACATCG
 CAAGGCTGTACCACCACGCTGACCCCGCAATCGACAAGGCTGCGGGATTTGATCT
 CTTCAAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGGAGAAA
 CTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCATCTGTGG
 GTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCTGCCTCCTTGTCTCAA
 CATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTACCATCGCTCTTCAGTCTTGTG
 ATTACCAGCGACTGCCCCACACATTCTCTGGGCCACCGGTTGAAAtctgglggttcttctggtg

gttctagcggcagcgagactcccggaaccagagtcggccacacccgaaagttccggaggagtagcggcgggTCTACA
CAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGAT
CCCACAGGGCAAGACCCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAG
GCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTA
TGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATC
GACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGG
CCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGAT
GCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAA
TGGCAAGGTGCTGAAGCAGCTGGGCACCCTGACCACAACCGAGCACGAGAACGCCCTG
CTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGAAC
GTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCCGCATCGTGCAGGACAACCT
CCCCAAGTTTAAGGAGAATTGTACATCTTCACACGCCCTGATCACCGCCGTGCCAGCC
TGCGGGAGCACCTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATC
GAGGAGGTGTTTTCTTCCCTTTTATAACCAAGCTGCTGACACAGACCCAGATCGACCTG
TATAACCAAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCCGAGAGAAGATCAAGGGCC
TGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCACATCATCGCC
TCCTGCCACACAGATTATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCCT
GTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCCTTCTGCAAGT
ACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAADGA
GCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCA
GCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAAT
CTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCCTG
AAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGAGGCAAGGAGCTGAGCG
AGGCCCTTCAAGCAGAAAACCGAGGATCCTGTCCACGCACACGCCGCCCTGGATCA
GCCACTGCCCTACAACCCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTG
GACAGCCTGCTGGGCCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACG
AGGTGAGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCCTC
TCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGA
AGTTCAAGCTGAACCTTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAG
AAGAACAATGGCGCCATCCTGTTTGTGAAGAAGGCCCTGTACTATCTGGGCATCATGCC
AAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCACAGAGAAAACCGAGGAG
GGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGC
AGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCATCCTGCT
GTCCAACAATTTATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCC
TGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAAG
GGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTA
TACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCT
GGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACCATCATGACTTCAGAGAATCG
CCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTAT
ACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGAC

CGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCC
 GAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGA
 AGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCGACACCCCTGTACCAG
 GAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCTGATGAGGCCAGGGC
 CCTGCTGCCCCAACGTGATCACCAAGGAGGTGTCTCAGGAGATCATCAAGGATAGGCGCT
 TTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATT
 CCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCTGAGACACCT
 ATCATCGGCATCGCCCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCAC
 CGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGA
 AGCTGGACAACAGGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGG
 GCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCTCCACGAGATCGTGGAG
 CTGATGATCCACTACCAGGCCGTGGTGGTGTGGGAGAACCCTGAATTTCCGCTTTAAGAG
 CAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACAGCAGTTTCGAGAAGATGCTGATC
 GATAAGCTGAATTGCCTGGTGTCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGC
 TGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCT
 GGCTTCCTGTTTTACGTGCCTGCCCATATACATCTAAGATCGATCCCTGACCGGCTTC
 GTGGACCCCTTCGTGTGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGA
 GGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACITTAAGAT
 GAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGGATATCG
 TGTTGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCTTTTCATCGCCGGCAAG
 AGAATCGTGCCAGTGATCGAGAATCACAGATTACCCGGCAGATACCGGGACCTGTATCC
 TGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCC
 AACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGTGGC
 CCTGATCCGCAGCGTGTCTGCAGATGCCGAACTCCAATGCCGCCACAGGCGAGGACTAT
 ATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCC
 AGAGTGGCCAATGGACGCCGATGCCAATGGCGCTACCACATCGCCCTGAAGGGCCAG
 CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAA
 TCAGGACTGGCTGGCTACATCCAGGAGCTGCGCAACaaaaggccggcgccacgaaaaaggccg
 gccaggcaaaaaagaaaaagGGATCCTCTGGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCT
GTGAGATATTATTGAAAAGGAGACCGGTAAAGCAACTGGTTATCCAGGAATCCATCCTCAT
GCTCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCCGGAAAGCGATATACTCGTG
CACACCGCTACGACGAGAGCACCGACGAGAATGTCATGCTTCTGACTAGCGACGCCC
CTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAAAG
ATGCTCTCTGGTGGTTCTcccaagaagaagaggaaagtc (SEQ ID NO:445)

Secuencia de aminoácidos de NLS(SV40)x2-rAPOBEC1-gsXTENgslinker-human-dAsCpf1(D908A)-
 NLS(nucleoplas- min)-GSlinker-UGI-NLS(SV40)

- 5 AsCpf1 en letra normal (AAs x-xxxx), rAPOBEC1 en **negrita** (AAs 23-250), NLS(nucleoplasmina) (krpaatkaggqakkkk, SEQ ID NO:7) en minúsculas, SV40 NLS en minúsculas (pkkrkv, SEQ ID NO:24), enlazador gsXTENgs (sggssggssgsetpgt- sesatpessggssgg, SEQ ID NO:25) en minúsculas, y UGI doble subrayado

MGpkkkrkvGSGpkkkrkvGSGSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCL
 LYEINWGGRRHSWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSWSPCGECSSRAI
 TEFLSRYPHYTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSN
 EAHWPYPHLLWVRLYVLELYCILGLPPCLNLRKQKQPLTFFIALQSCHYQRLPPHILWAT
 GLKsggsgsgsgsgselpgtssesatpassggsgsgSTQFEGFTNLYQVSKTLRFELIPQGGKTLKHIEQGGFI
 EEDKARNHDHYKELKPIIDRIYKYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNLIEEQAT
 YRNAIHDIYFIGRTDNLDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKF
 TTYFSGFYENRKNVSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIG
 IFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIA
 SLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCYKTLNRNENVLETAELFNEINSDLT
 HIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQSRSLKHEDINLQEIISAAG
 KELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFAVDESNE
 VDPEFSARLTGKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNN
 GAILFVKNGLYLYGIMPKQKGRYKALSFEPTKTESEGFDMYYDYFPDAAKMIPKCTSQLKAV
 TAHFQTHHTPILLSNMFIEPLEITKEIYDLNNPEKEPKKFQAYAKKTGDQKGYREALCKWIDF
 TRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLF
 QIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGE
 KMLNKKLDQKTPIDTLYQELYDYVNHRLSHDLSEARALLPNVITKEVSHEIKDRRFTSDK
 FFFHVPITLNYQAANSFKFNQVRNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTI
 QQFDYQKKLDNREKERVAAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNF
 GFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGLNPNYQLTDQFTSFAMGTQS
 GFLFYVPAPYTSKIDPLTGFDVDFVWTKIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRN
 LSFQRLPGFMPAWDIVFEKNETQFQDAKGTPIAGKRIVPVIEHNRFTGRYRDLYPANELIAL
 EEKGIVFRDGSNILPKLLENDSDHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCF
 DSRFQNPWFPMADANGAYHIALKGQQLLNHLKESKDLKLQNGISNQDWLAYIQELRNkrpaa
 tkkagqakkkkGSSGGSGSGSGGSTNLSDIIEKETGKQLVQESILMLPEEVFEVIGNKPESDILVHT
AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSpkkkrkv (SEQ ID NO:446)

Secuencia de nucleótidos de (RTW1028) pCAG-NLS(SV40)x2-rAPOBEC1-gsXTENgslinker-human-dAsCpf1 (D908A)triple- variante(E174R/S542R/K548R)-NLS(nucleoplasmina)-GSlinker-UGI-NLS(SV40)

- 5 Codón humano optimizado dAsCpf1 (D908A) en letra normal (NTs 844-4764), rAPOBEC1 en **negrita** (NTs 67-750), Nucleoplasmina NLS en minúsculas

(aaaagggccggccggccacgaaaaaggccggccaggcaaaaaagaaaaag, SEQ ID NO:21), SV40 NLS en minúsculas (ccaaa-gaaaaagaggaaagtc, cctaaaaagaaacgaaagggt, o cccaagaagaagaggaaagtc, SEQ ID NOs:19, 20, o 22, respectivamente), enlazador gsXTENgs

- 10 (tctggtggttcttctggtggttagcggcagcgagactcccgggacctcagagtcgccacacccgaaagttccggagggagtag cgggcggg, SEQ ID NO:23) en minúsculas, y UGI doble subrayado

ATGGGCCCCaaagaaaaagaggaagtcGGCAGTGGACtaaaaagaaaagaaaggttGGGTGAG
 GTAGCTCAGAGACTGGCCAGTGGCTGTGGACCCACATTGAGACGGCGGATCGAGC
 CCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCT
 TTACGAAATTAATTGGGGGGGCGGCACTCCATTTGGCGACATACATCACAGAACACT
 AACAAGCACGCTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTC
 CGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAG
 TAGGGCCATCACTGAATTCTGTCAAGGTATCCCCACGTCACCTGTTTATTTACATCG
 CAAGGCTGTACCACCACGCTGACCCCGCAATCGACAAGGCCTGCGGGATTTGATCT
 CTTCAAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGGAGAAA
 CTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCTGTGG
 GTACGACTGTACGTTCTTGAAGTGTACTGCATCATACTGGGCCTGCCTCCTTGTCTCAA
 CATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTACCATCGCTCTTCACTCTTGTCT
 ATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAAtctgggtgttcttgggtg
 gttctagcggcagcgagactcccggaacctcagagtcggccacacccgaagtcgggagggagtagcgggggTCTACA
 CAGTTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGAT
 CCCACAGGGCAAGACCCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAG
 GCCCCGAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTA
 TGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCGGCCATC
 GACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGG
 CCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGAT
 GCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAA
 TGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCTG
 CTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGCCAACAGGAAGAAC
 GTGTTTCAGCGCCGAGGATATCAGCACAGCCATCCACACCGCATCGTGACAGGACAACCT
 CCCCAGTTTAAAGGAGAAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCC
 TGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCTGTGAGCACCTCCATC
 GAGGAGGTGTTTTCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTG
 TATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCC
 TGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCACATCATCGCC
 TCCCTGCCACACAGATTCTATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCCT
 GTCTTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCCTTCTGCAAGT
 ACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGA
 GCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCA
 GCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAAT
 CTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTG
 AAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGAGGCAAGGAGCTGAGCG
 AGGCCTTCAAGCAGAAAACAGCGAGATCCTGTCCACGCACACGCCGCCCTGGATCA

GCCACTGCCCTACAACCCCTGAAGAACGACGAGGAGAGAGATCCTGAAGTCTCAGCTG
 GACAGCCTGCTGGCCCTGTACCACCTGCTGGACTGGTTTGGCGTGGATGAGTCCAACG
 AGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCCTT
 TCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGA
 AGTTCAAGCTGAACCTTTCAGATGCCTACACTGGCCGCCGGCTGGGACGTGAATAAGGCC
 AAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCC
 AAAGCAGAAGGGGACGATATAAGGCCCTGAGCTTCGAGCCACACAGAAAAACCAGCGAG
 GGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGC
 AGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCATCCTGCT
 GTCCAACAATTTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCC
 TGAGAAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAAACCGGCCACCAAG
 GGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTA
 TACCAAGACAACCTCTATCGATCTGTCTAGCCTGCCGCCATCCTCTCAGTATAAGGACCT
 GGGCGAGTACTATGCCGAGCTGAATCCCTGCTGTACACATCAGCTTCCAGAGAATCG
 CCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTAT
 AACAAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGAC
 CGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCC
 GAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGA
 AGATGCTGAACAAGAAGCTGAAGGATCAGAAAAACCCCAATCCCCGACACCCTGTACCAG
 GAGCTGTACGACTATGTGAATCACAGACTGTCCACGACCTGTCTGATGAGGCCAGGGC
 CCTGCTGCCAACGTGATACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCT
 TTACCAGCGACAAGTTCTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATT
 CCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCT
 ATCATCGGCATCGCCCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCAC
 CGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGA
 AGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGG
 GCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCTATCCACGAGATCGTGGAC
 CTGATGATCCACTACCAGGCCGTGGTGGTGTGGAGAACCTGAATTTCCGGCTTTAAGAG
 CAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAAGCAGTTTCGAGAAGATGCTGATC
 GATAAGCTGAATTGCCTGGTGTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCCGTGC
 TGAACCCATACCAAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCT
 GGCTTCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCTGACCGGCTTC
 GTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCTTGGGA
 GGGCTTGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGAT
 GAACAGAAATCTGTCTTCCAGAGGGGCTGCCCGGCTTTATGCCTGCATGGGATATCG
 TGTTGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAG
 AGAATCGTGCCAGTGATCGAGAATCACAGATTCACGGGCAGATACCGGGACCTGTATCC
 TGCCAACGAGCTGATCGCCTGCTGGAGGAGAAGGGCATCGTGTTACGGGATGGCTCC
 AACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGTGGC
 CCTGATCCGACGCTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTAT

ATCAACAGCCCCGTGCGGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCC
 AGAGTGGCCAATGGACGCCGATGCCAATGGCGCTACACATCGCCCTGAAGGGCCAG
 CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAA
 TCAGGACTGGCTGGCTACATCCAGGAGCTGCGCAACaaaaggccggccgcaaaaaaggccg
 gccaggcaaaaaagaaaaagGGATCCTCTGGTGTTCTGGAGGATCTGGTGTTCTACTAATCT
GTCAGATATTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCAT
GCTCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCCGGAAGCGATATACTCGTG
CACACCGCTACGACGAGAGCACCGACGAGAATGICATGCTTCTGACTAGCGACGCC
CTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAA
ATGCTCTCTGGTGTTCTTccaagaagaagaggaaagtc (SEQ ID NO:447)

Secuencia de aminoácidos de NLS(SV40)x2-rAPOBEC1-gsXTENgslinker-human-dAsCpf1(D908A)
 triplevariante(E174R/S542R/K548R)-NLS(nucleoplasmina)-GSlinker-UGI-NLS(SV40)

- 5 AsCpf1 en letra normal (AAs x-xxxx), rAPOBEC1 en **negrita** (AAs 23-250), NLS(nucleoplasmina) (krpaatkagqakkkk, SEQ ID NO:7) en minúsculas, SV40 NLS en minúsculas (pkkkrv, SEQ ID NO:24), enlazador gsXTENgsl (sggssggsgsetpgt- sesatpessggsgg, SEQ ID NO:25) en minúsculas, y UGI doble subrayado

MGpkkkrvGSgpkkkrvGSgSSEtGPVAVDPTLRRRIEPHEFEVFFDPRELKETCL
 LYEINWGGRRHSWVRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPGCECSRAI
TEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSN
EAHWPRYPHLWVRLYVLELYCILGLPPCLNLRKQPKLTFFIALQSCHYQRLPPLHWAT
 GLKsggssggsgsetpgtsesatpessggsggSTQFEGFTNLYQVSKTLRFELIPQGKTLKHIEQGF
 EEDKARNHDHYKELKPIIDRIYKYADQCLQLVQLDWNLSAIDSYRKEKTEETRNALIEEQAT
 YRNAIHDFYIGRTDNLTDANKRHAIEYKGLFKAELFNGKVLKQLGTVTTEHENALLRSFDKF
 TTYFSGFYRNKRVFSAEDISTAIPHRIVQDNFPKFENCHIFTRLITAVPSLREHFENVKKAIG
 IFVSTSIIEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIA
 SLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCYKTLRNENVLETAELFNLNSIDL
 HIFISHKKLETISSALCDHWDLRNALYERRISELTGKITSAKEKVQSRSLKHEDINLQEIISAAG
 KELSEAFKQKTSEILSHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNE
 VDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFLNFQMPFLARGWQVNRKNN
 GAILFVNGLYLYLGIMPKQKGRYKALSFEPTTEKTSSEGFDMYDYFPDAKMIPKCSQTLKAV
 TARFQTHHTPILLSNNFIEPLEITKEIYDLNPEKEPKKFQAYAKKTGDQKGYREALCKWIDF
 TRDFLSKYTKTSSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLF
 QIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGE
 KMLNKKLDQKTPIDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIKDRRFTSDK
 FFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLYITVIDSTGKILEQRSNTI
 QQFDYQKKLDNREKERVAAARQAWSVVGTIKDLKQGYLSQVIHEIVOLMIHYQAVVLENLNF
 GFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGLNPNYQLTDQFTSFAKMGTSQ
 GFLFYVPAPYTSKIDPLTGFDVDFVWTKNHNESRKHFLGFDLHYDVKTGDFILHFKMNRN
 LSFQRLGPGFMPAWDIVFEKNETQFDAQGTPIAGKRIVPIENHRTGGRYRDLYPANELIAL
 EEKGIVFRDGSNILPKLLENDDSHAITMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCF
 DSRFQNPWPMDADANGAYHIALKGQLLNHLKESKDLKLQNGISNQDWLAYIQELRNkrpaa
lkagqakkkkGSSGSGSGSGSTNLSDIIEKETGKQLVQESILMLPEEVEEVIGNKPESDILVHT
AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSpkkkrv (SEQ ID NO:448)

Ejemplo 1. Variantes de AsCpf1 con especificidad PAM alterada

- 10 Para intentar alterar el rango de selección de las nucleasas Cpf1, primero examinamos las estructuras cristalinas disponibles de AsCpf1 y LbCpf1 (Dong, Nature 2016; Yamano, Cell 2016). Entre otras observaciones, estas

estructuras demuestran que la especificidad PAM está mediada por una combinación de interacciones electrostáticas y lectura indirecta de bases. Por tanto, nuestra hipótesis es que determinadas combinaciones de sustituciones de aminoácidos en residuos en estrecha proximidad espacial a las bases de ADN de las bases de PAM podrían dar lugar a variantes con preferencias de reconocimiento de PAM modificadas o relajadas. Para probarlo, examinamos regiones de AsCpf1 en las proximidades de los PAM que abarcan residuos G131-L137, S161-S181, N534-I555, Y595- T616, L628-F632, y S685-I693 (Tabla 1). Nos centramos en los aminoácidos de la secuencia de referencia de AsCpf1 cuya posición tridimensional cumplía al menos uno de los siguientes criterios: 1) proximidad espacial a bases de ADN PAM (tanto en la cadena diana como en la no diana), 2) posicionamiento dentro del surco mayor o menor del ADN, y/o 3) residuos posicionados de tal manera que la sustitución del aminoácido existente por una alternativa cargada positivamente, como arginina, lisina o histidina, podría aumentar la proximidad (y presumiblemente la interacción) de la cadena lateral con la estructura del fosfodiéster. Dado que las estructuras cristalinas que contienen el ARNcr y el ADN que contiene PAM sólo están disponibles para AsCpf1, se identificaron 25 posiciones homólogas en LbCpf1 y FnCpf1 basadas en la alineación de secuencias (Tabla 1) entre los tres ortólogos.

Tabla 1: Comparación de residuos candidatos a mutación de AsCpf1 y LbCpf1 para crear variantes de especificidad de reconocimiento de PAM alteradas. Las alineaciones se realizaron con o sin FnCpf1.

AsCpf1	LbCpf1	LbCpf1 (+18)	FnCpf1	parámetros de alineación	
G131-L137	G131	S117*	S135*	G133**	* significa residuos de LbCpf1, sólo de la alineación con AsCpf1
	L132	L118*	L136*	L142**	** significa residuos de FnCp-F1, sólo de la alineación con AsCpf1
	F133	F119*	F137*	K143**	significa residuos de la alineación con los 3 ortólogos de Cpf1 (AsCpf1 como referencia)
	K134	K120*	K138*	Q144**	
	A135	K121*	K139*	S145**	
	E136	D122*	D140*	K146**	
	L137	I123*	I141*	D147**	
5161-5181	S161	5143	S161	s171	TEXTO negro significa residuos de la alineación con los 3 ortólogos de Cpf1
	F162	F144	F162	F172	
	D163	N145	N163	K173	
	K164	G146	G164	G174	
	F165	F147	F165	W175	
	T166	T148	T166	T176	
	T167	T149	T167	T177	
	Y168	A150	A168	Y178	
	F169	F151	F169	F179	
	5170	T152	T170	K180	
	G171	G153	G171	G181	
	F172	F154	F172	F182	
	Y173	F155	F173	H183	

(continúa)

AsCpf1	LbCpf1	LbCpf1 (+18)	FnCpf1	parámetros de alineación
E174	D156	D174	E184	
N175	N157	N175	N185	

ES 2 992 912 T3

	R176	R158	R176	R186	
	K177	E159	E177	K187	
	N178	N160	N178	N188	
	V179	M161	M179	V189	
	F180	F162	F180	Y190	
	5181	5163	5181	5191	
N534-1555	N534	Y524	Y542	N599	TEXTO negro significa residuos de la alineación con los 3 ortólogos de Cpf1
	F535	F525	F543	F600	
	Q536	Q526	Q544	E601	
	M537	N527	N545	N602	
	P538	P528	P546	S603	
	T539	Q529	Q547	T604	
	L540	F530	F548	L605	
	A541	M531	M549	A606	
	5542	G532	G550	N607	
	G543	G533	G551	G608	
	W544	W534	W552	W609	
	D545	D535	D553	D610	
	V546	K536	K554	K611	
	N547	D537	D555	N612	
	K548	K538	K556	K613	
	E549	E539	E557	E614	
	K550	T540	T558	P615	
	N551	D541	D559	D616	
	N552	Y542	Y560	N617	
	G553	R543	R561	T618	
	A554	A544	A562	A619	
	1555	T545	T563	1620	
Y595-T616	Y595	Y583	Y601	Y659	TEXTO negro significa residuos de la alineación con los 3 ortólogos de Cpf1
	D596	K584	K602	K660	
	Y597	L585	L603	L661	
	F598	L586	L604	L662	
	P599	P587	P605	P663	
	D600	G588	G606	G664	
	A601	P589	P607	A665	
	A602	N590	N608	N666	
	K603	K591	K609	K667	
	M604	M592	M610	M668	
	1605	L593	L611	L669	
	P606	P594	P612	P670	
	K607	K595	K613	K671	
	C608	V596	V614	V672	
	5609	F597	F615	F673	
	T610	F598	F616	F674	
	Q611	5599	5617	5675	

(continúa)

	AsCpf1	LbCpf1	LbCpf1 (+18)	FnCpf1	parámetros de alineación
	L612	K600	K618	A676	
	K613	K601	K619	K677	
	A614	W602	W620	5678	
	V615	M603	M621	1679	
	T616	A604	A622	K680	
L628-F632	L628	F598*	F616*	R690**	* significa residuos de LbCpf1, sólo de la alineación con AsCpf1 ** significa residuos de FnCpf1, sólo de la alineación con AsCpf1 si la alineación se realiza con AsCpf1, LbCpf1, y FnCpf1, tanto LbCpf1 y FnCpf1 no se alinean en esta región con AsCpf1
	5629	5599*	S617*	1691**	
	N630	K600*	K618*	R692**	
	N631	K601*	K619*	N693**	
	F632	W602*	W620*	H694**	
5685-1693	5685	5644	5662	5729	TEXTO negro significa residuos de la alineación con los 3 ortólogos de Cpf1 ** significa residuos de FnCpf1, sólo de la alineación con AsCpf1
	K686	R645	R663	K730	
	Y687	Y646	Y664	H731	
	T688	P647	P665	P732	
	K689	K648	K666	E733	
	T690	W649	W667	W734	
	T691	5650	5668	K735	
	5692	N651	N669	D736	
	1693	A652	A670	F737**	

En los experimentos iniciales, primero tratamos de definir cuidadosamente las preferencias de PAM de AsCpf1 y LbCpf1 de tipo salvaje probando las actividades de estas nucleasas en células humanas contra sitios diana PAM alternativos que tienen diferencias de base dentro del motivo TTTN definido en la caracterización inicial de estas enzimas (*Zetsche, Cell 2015*). Usando nuestro reconocido ensayo de disrupción de EGFP en células humanas, primero probamos las capacidades de AsCpf1 y LbCpf1 para mutagenizar varios sitios diana que albergaban PAMs TTTN canónicos o PAMs no canónicos presentando una serie de discordancias de una sola base dentro del motivo TTT. Determinamos que aunque tanto AsCpf1 como LbCpf1 podían tolerar bases no canónicas en los PAMs, el reconocimiento por parte de LbCpf1 era más promiscuo (Fig. 2). Ambas enzimas Cpf1 se dirigieron de forma eficiente y consistente a los sitios en nuestro ensayo de disrupción de EGFP con PAMs CTTN, GTTN y TTCN alternativos. (El ensayo de disrupción de EGFP mide la pérdida de expresión de EGFP como un sustituto para la selección de sitios dentro de la secuencia de EGFP por nucleasas de edición del genoma (*Reyon, Nature Biotechnology 2012*)). También examinamos la capacidad de AsCpf1 para reconocer sitios que contienen secuencias PAM más divergentes en EGFP, con dos o tres bases de diferencia en la secuencia TTT de un PAM TTTN. Con la excepción de cierta actividad leve y variable contra sitios que contienen PAMs CCCN, CCTN, GATN, GCTN y TCCN, descubrimos que la nucleasa AsCpf1 de tipo salvaje no se dirigía eficazmente a ninguno de estos PAMs alternativos (Fig. 3).

Sin embargo, como la actividad observada en el ensayo de disrupción de EGFP podría representar una combinación de disrupción génica mediada por nucleasas y represión transcripcional mediada por la unión al ADN (sin escisión), probamos las actividades de AsCpf1 y LbCpf1 en sitios diana de genes humanos endógenos que contienen secuencias PAM alternativas con diferencias de una sola base en relación con los PAM canónicos, porque la lectura de este ensayo cuantifica los eventos de disrupción génica *bona fide* (Fig. 4). Descubrimos que, aunque se observaron actividades modestas con AsCpf1 y LbCpf1 en sitios que albergaban ciertos sitios PAM no canónicos

en nuestro ensayo de disrupción de EGFP, no observamos una inducción comparable de mutaciones indel en sitios de genes endógenos que portaban estas secuencias PAM alternativas. Este resultado sugiere que, aunque las nucleasas AsCpfI y LbCpfI de tipo salvaje pueden unirse eficazmente a secuencias alternativas de PAM con diferencias de una sola base, en algunos casos pueden no escindir eficazmente estos sitios.

- 5 Dada la limitada capacidad de las nucleasas CpfI de tipo salvaje para escindir PAMs no-TTTN, intentamos diseñar variantes de CpfI que pudieran dirigirse e interrumpir sitios que albergaran dichos PAMs alternativos. En los experimentos iniciales, intentamos diseñar nucleasas CpfI con especificidades de reconocimiento de PAM relajadas. Utilizamos el ensayo de disrupción de EGFP para identificar y analizar rápidamente sustituciones de aminoácidos en AsCpfI que pudieran alterar el reconocimiento de PAM, teniendo en cuenta que, dadas las limitaciones del
- 10 ensayo basado en EGFP descrito anteriormente, tendríamos que validar en última instancia cualquier variante que identificáramos por su capacidad para escindir sitios dentro de genes endógenos reales en células humanas. Para probar la hipótesis de si las sustituciones racionales de AsCpfI podrían alterar la especificidad de PAM, nos centramos en probar el efecto de las sustituciones de aminoácidos en las posiciones: T167, S170, E174, T539, S542, K548, N551, N552, M604, y K607.
- 15 Nótese que la numeración posterior de PAM se basa en que el PAM TTTN se numera $T_4T_3T_2N_1$, y que sólo se examinó un pequeño subconjunto de todos los PAMs posibles para las pruebas iniciales de cada variante de AsCpfI, con pruebas posteriores más amplias para combinaciones de variantes.

T167 / T539: Basado en la proximidad de T539 a T_4 del PAM en la estructura de co-cristal de AsCpfI, previmos que las sustituciones concurrentes de T539K/T167A o T539R/T167A podrían permitir el reconocimiento específico de base de una G en la cuarta posición del PAM por: 1) la sustitución T539 en K o R que permite la lectura específica de base de una guanina, y/o 2) la sustitución T167A que atenúa otros contactos interferentes o desfavorables inducidos por la alteración del residuo T539. En relación con AsCpfI de tipo salvaje, tanto la variante T539K/T167A como la T539R/T167A muestran una actividad mejorada en el ensayo de disrupción de EGFP en sitios que presentan PAMs GTTN con reducciones mínimas o sólo modestas de la actividad en sitios que presentan PAMs

20 TTTN canónicos (Fig. 5A).

S170 / E174: Ambos residuos se encuentran en el surco principal del ADN, con S170 cerca de T_2 del PAM y E174 cerca de la posición T_2 o N_1 del PAM y cerca de la cadena diana de la estructura del ADN. Entendemos que las sustituciones de arginina en estas posiciones podrían relajar la especificidad de PAM, permitiendo la creación de contactos no específicos con la estructura del ADN o estableciendo potencialmente el reconocimiento específico de bases de PAMs TTGN o TTTG. Tanto la variante S170R como la E174R aumentan la actividad en los PAMs

30 TTTN canónicos en el ensayo de disrupción de EGFP, a la vez que aumentan la actividad en los sitios con PAMs GTTN (Fig. 5B). También mostramos que tanto las variantes S170R como E174R confieren una mayor capacidad para dirigirse a sitios Pam CTTN y TTCN (Fig. 8C).

S542: Este residuo está situado en el surco principal, muy cerca de las bases T_3 y T_2 de PAM. Por tanto, planteamos la hipótesis de que las mutaciones S542K o S542R podrían funcionar para relajar la especificidad de PAM mediante: 1) proporcionar energía no específica adicional en la interfaz de unión de PAM para acomodar bases no canónicas, y/o 2) crear un contacto de base específica que pudiera reconocer potencialmente un G_3 o G_2 en la cadena no diana, o quizás un G_3 o G_2 en la cadena diana que sería un C_3 o C_2 en la cadena no diana de PAM. Nuestras hipótesis predicen que las variantes con sustituciones en S542 podrían permitir el reconocimiento de sitios que albergan PAMs TGTN, TTGN, TCTN o TTCN. Utilizando el ensayo reporter EGFP, descubrimos que las variantes de AsCpfI con una sustitución S542K o S542R (pero no con una sustitución S542Q) muestran un aumento de la actividad en sitios diana con PAMs TTCN no canónicos (Fig. 5C). Curiosamente, cuando se probaron sus capacidades para escindir y mutagenizar dianas génicas humanas endógenas, las variantes de AsCpfI que portaban una mutación S542Q, S542K o S542R mostraron mayores capacidades para inducir mutaciones indel en

40 sitios diana que portaban PAMs TTTN (Fig. 5D), pero sólo las variantes S542K y S542R mostraron mayores actividades en sitios diana que portaban PAMs TTCN no canónicos (Fig. 5D). El reconocimiento de PAMs no canónicos adicionales por estas variantes también se examina con más detalle en la Fig. 7 (ver más adelante).

K548: Este residuo está situado cerca de A_4 y A_3 de la cadena de ADN no PAM y cerca de la estructura del ADN de la cadena diana. Por tanto, nuestra hipótesis es que las sustituciones en esta posición podrían aumentar la actividad frente a sitios diana con PAMs CTTN, TCTN o CCTN no canónicos. Descubrimos que la introducción de una mutación K548R no parece conferir ninguna alteración sustancial en la especificidad de PAM por sí sola, pero sí contribuye a relajar el reconocimiento de PAM en el contexto de otras sustituciones en las posiciones S542, N551 y N552 (ver más adelante en la Fig. 9).

50

N551 / N552: Los residuos N551 y N552 están situados en el surco principal entre las estructuras de las cadenas de ADN diana y no diana, pero N552 también está muy cerca de A_3 de la cadena de ADN no PAM y cerca de la estructura de ADN de la cadena diana. Mientras que la sustitución N551R no parece tener ningún efecto perjudicial

55

o, en un caso, quizás mejore ligeramente la actividad de AsCpfl en sitios diana con secuencias de PAMs GTTN no canónicas (sin afectar al reconocimiento de sitios con PAMs TTTN canónicos), la sustitución N552R parece anular la actividad en sitios diana ya sea con PAMs TTTN o GTTN (Fig. 5E). También exploramos las sustituciones N551R y N552R en combinación con la mutación S542R y otras combinaciones de mutaciones (ver Fig. 8B y Fig. 11 más abajo).

M604 / K607: El residuo M604 se posiciona en el surco menor del ADN cerca de la posición N₁ del PAM y del nucleótido A₂ de la cadena de ADN diana no PAM. El residuo K607 también está situado en el surco menor y forma una red con T₃ y T₂ (de PAM) y A₃ de la cadena de ADN no PAM. Múltiples sustituciones diferentes en K607 por sí solas parecen tener un impacto negativo en la actividad de AsCpfl (Fig. 5f), pero la combinación de una sustitución K607H junto con la mutación S542R conduce a una variante con mayor actividad contra sitios que presentan PAMs TTTN canónicos o TTCN no canónicos (según lo juzgado por el ensayo de disrupción de EGFP o por la mutagénesis de sitios diana de genes humanos endógenos, en las Figs. 5g y 5h, respectivamente). De forma similar, una sustitución M604A combinada con una sustitución S542R mejora la actividad frente a sitios diana que presentan PAMs TTTN canónicos o TTCN no canónicos cuando se ensaya utilizando el ensayo de disrupción EGFP (Fig. 5g).

Para seguir probando el fenotipo de especificidad de PAM relajada resultante de la introducción de la mutación S542R, comparamos la actividad de disrupción de EGFP de AsCpfl de tipo salvaje con la de AsCpf1-S542R en sitios diana de EGFP que presentan un PAM con una única diferencia de base con respecto a la secuencia canónica (incluyendo los cuatro PAMs no canónicos que, según nuestra hipótesis, podrían ser reconocidos por la variante S542R) (Fig. 6A). En estos experimentos, observamos un aumento de la actividad de la variante AsCpf1-S542R frente a sitios diana con múltiples sitios PAM no canónicos, incluyendo PAMs más allá de los cuatro hipotetizados por nuestro modelo de contacto base-específico (lo que sugiere una mejora general en la afinidad de unión del PAM y una relajación relacionada en la especificidad del PAM). Para examinar más a fondo el potencial del mutante S542 para ampliar el rango de selección de AsCpfl, comparamos el AsCpfl de tipo salvaje con la variante S542R en una serie de sitios diana de EGFP con PAMs que presentan 2 o 3 diferencias de base en relación con el sitio canónico (Fig. 6B). El mutante S542R mantuvo al menos el mismo nivel de actividad observado con el AsCpfl de tipo salvaje en todos los sitios, y mejoró drásticamente la actividad (en algunos casos varias veces) frente a muchos sitios con PAMs que presentaban dos o tres sustituciones (Fig. 6B). Nuestros resultados con el ensayo de disrupción de EGFP sugieren que la variante AsCpf1-S542R puede reconocer sitios que presentan los siguientes PAMs: TTTN, CTTN, GTTN, TCTN, TGTN, TTAN, TTCN, ATCN, CCCN, CCTN, GCTN, GGTN, TCCN, y TGCN. A continuación probamos la variante S542R en sitios diana de genes humanos endógenos con PAMs con una (Fig. 7A) y 2 o 3 sustituciones de bases (Fig. 7B). Estos experimentos revelaron de nuevo que la variante S542R puede escindir una gama más amplia de motivos PAM discordantes, pero el espectro de estos sitios no fue tan amplio como el que observamos en el ensayo de disrupción de EGFP. Basándonos en los resultados de nuestros experimentos con estos sitios diana de genes humanos endógenos, concluimos que la variante AsCpf1-S542R puede escindir sitios que presentan los siguientes PAMs: TTTN, ATTN, CTTN, GTTN, TCTN, TTCN, CCCN, y TCCN.

La observación de que una única sustitución en S542 podía ampliar la especificidad de reconocimiento de PAM de AsCpfl lo que sugiere que podría ser posible aumentar aún más el rango de selección añadiendo mutaciones únicas o múltiples a esta variante. Como se muestra en la Figura 5 y se ha descrito con detalle más arriba, descubrimos que las sustituciones de aminoácidos en S170, E174, K548, N551 y K607 (solos o combinadas) daban lugar a especificidades de reconocimiento de PAM algo alteradas. Así pues, intentamos explorar si otras combinaciones de sustituciones en estas posiciones junto con la mutación S542R podrían mejorar aún más el rango de selección de AsCpfl. En primer lugar, determinamos que las combinaciones de sustituciones que incluían S542R/K548R, S542R/N551R y K548R/N551R podían, en la mayoría de los casos, mejorar la actividad con respecto a la sustitución S542R sola en los sitios diana que presentaban PAMs TTTN canónicos o GTTN no canónicos (Fig. 8A). A continuación, utilizando el ensayo de disrupción de EGFP, determinamos a través de un mayor número de sitios diana con secuencias PAM más diversas: 1) que las variantes que presentan las sustituciones únicas S170R o E174R podrían superar en muchos sitios diana a la sustitución S542R, 2) las variantes E174R/S542R, S542R/K548R y S542R/NSS 1R funcionan tan bien o mejor que la S542R sola en una serie de sitios diana diferentes, y 3) que la variante de triple sustitución E174R/S542R/K548R presentaba el mayor nivel de actividad entre una amplia serie de variantes de AsCpfl que probamos contra una serie de sitios diana que presentaban sitios PAM TTTN canónicos y CTTN, GTTN, TATN, TCTN, TGTN, TTAN, TTCN y TTGN no canónicos (Figs. 8B y 8C).

Otra comparación más de la variante AsCpfl E174R/S542R/K548R con la variante AsCpfl de tipo salvaje y la variante AsCpf1-S542R en los sitios diana de EGFP con PAMs con diferencias de una sola base (Fig. 9A) reveló que la variante E174R/S542R/K548R tenía una mayor actividad de disrupción de EGFP que la AsCpfl de tipo salvaje y que la S542 en casi todos los sitios diana con varios PAM ensayados. Por otra parte, cuando la variante E174R/S542R/K548R se comparó con la AsCpfl de tipo salvaje en una serie de sitios EGFP con PAMs que presentaban dos o tres discordancias, se observaron aumentos sustanciales en la disrupción de EGFP en muchos de estos sitios (Fig. 9B). La variante de triple sustitución E174R/S542R/K548R también mostró actividades

generalmente superiores y en una gama más amplia de PAMs variantes que la variante S542R (comparar la Fig. 9B con la Fig. 6B), incluyendo el reconocimiento eficiente de sitios que presentan los siguientes PAMs: TTTN, CTTN, GTTN, TATN, TCTN, TGTN, TTAN, TTCN, TGTN, AGTN, ATCN, CATN, CCCN, CCTN, CGTN, CTAN, CTCN, GATN, GCTN, GGTN, GTCN, TACN, TCCN, y TGCN.

Debido a la limitación del ensayo de disrupción de EGFP mencionado antes para evaluar las actividades de las nucleasas CpfI, ahora evaluamos la actividad de nuestra variante AsCpfI E174R/S542R/K548R frente a una serie de sitios diana de genes humanos endógenos que presentan PAMs con diferencias de una sola base (Fig. 10A) o de dos o tres bases (Fig. 10B). Estos resultados demuestran que AsCpfI E174R/S542R/K548R puede escindir eficazmente sitios con los siguientes PAMs: TTTN, ATTN, CTTN, GTTN, TATN, TCTN, TGTN, TTCN, ATCN, CCCN, CCTN, CTCN, GCTN, GGTN, TCCN, y TGCN. Es importante señalar que AsCpfI-E174R/S542R/K548R no se probó en sitios diana con todos los PAMs posibles. Significativamente, los sitios diana que presentaban la mayoría de estos PAMs alternativos no se pudieron escindir ni siquiera con LbCpfI de tipo salvaje, que tiene una preferencia de PAM más relajada que AsCpfI de tipo salvaje (Figs. 2 y 4). Por otra parte, para los sitios PAM variantes que podrían ser escindidos por LbCpfI de tipo salvaje, observamos que la variante AsCpfI-E174R/S542R/K548R superó sistemáticamente a LbCpfI de tipo salvaje a juzgar por la eficiencia de la inducción de mutación indel. Curiosamente, la variante AsCpfI-E174R/S542R/K548R también mostró una actividad mejorada frente a los sitios PAM TTTN canónicos, demostrando incluso una actividad sustancial frente al sitio TTTN-6, previamente no seleccionable con AsCpfI o LbCpfI (Fig. 10A y 10B). El sitio TTTN-6 lleva una T en la primera posición del PAM (para un PAM TTTT), lo que sugiere que la variante AsCpfI de triple sustitución puede mejorar la actividad frente a sitios con una T en la primera posición del PAM. Así pues, la variante AsCpfI-E174R/S542R/K548R mejora sustancialmente el rango de selección de la plataforma CpfI para sitios con PAMs no canónicos en relación con las nucleasas AsCpfI y LbCpfI de tipo salvaje y, en general, también muestran actividad mejorada en sitios con PAMs canónicos.

A continuación, para intentar relajar aún más la especificidad PAM de nuestras variantes PAM AsCpfI y/o mejorar la magnitud de la actividad en cualquier PAM determinado, añadimos más sustituciones de aminoácidos a la variante E174R/S542R/K548R. Primero, añadimos la sustitución N551R o N552R para generar las variantes de sustitución cuádruple E174R/S542R/K548R/N551R y E174R/S542R/K548R/N552R, respectivamente. La comparación de estas dos variantes de sustitución cuádruple con AsCpfI de tipo salvaje y la variante E174R/S542R/K548R reveló que la variante E174R/S542R/K548R/NSS 1R podía mejorar la actividad de disrupción génica en sitios que albergaban diversos PAM no canónicos que presentaban diferencias simples, dobles o triples, a juzgar tanto por el ensayo de disrupción EGFP como por sus capacidades para inducir mutaciones indel en sitios diana de genes humanos endógenos (Figs. 11A y 11B, respectivamente). Por el contrario, la variante de sustitución cuádruple E174R/S542R/K548R/N552R no mostró una mejora de la actividad en estos mismos experimentos y, en muchos casos, incluso la anuló (Figs. 11A y 11B). Para seguir comparando las especificidades de reconocimiento de PAM de las variantes E174R/S542R/K548R y E174R/S542R/K548R/N551R, examinamos sus actividades en un mayor número de sitios diana de genes humanos endógenos que contenían PAMs canónicos o no canónicos (diferencia de una sola base). Observamos una actividad comparable de ambas variantes en la mayoría de los sitios, con un pequeño número de casos en los que una u otra variante mostró una actividad ligeramente mejorada (Fig. 11C).

Ejemplo 1B. Caracterización adicional de variantes de AsCas12a con especificidades PAM alteradas y actividades mejoradas *on-target*

Caracterizaciones previas de ortólogos de Cas12a en células humanas revelaron que As y LbCas12a eran nucleasas consistentemente más efectivas en sitios con PAMs TTTV (Kim et al., Nat Biotechnol., 2016, 34:863-8), y que Fn y MbCas12a pueden poseer preferencias de PAM relajadas de NTTN (Zetsche et al., Cell, 2015, 163:759-71). Para evaluar con más detalle las actividades y preferencias de PAM de cada ortólogo, se examinaron en células humanas sus actividades de edición del genoma utilizando dos conjuntos de doce ARNcr dirigidos a sitios que albergaban PAMs TTTN o VTTN (Fig. 19A). Se observó una interrupción génica similar entre los cuatro ortólogos en los sitios PAM TTTN, aunque se observaron diferencias específicas. Por otra parte, Fn y Mb podían dirigirse más eficazmente a los PAMs VTTN en comparación con As y LbCas12a, pero en consonancia con informes previos sus actividades medias en los sitios VTTN eran demasiado bajas para caracterizar estos PAMs como dianas *bona fide* (Figs. 19A y 19B). Estos resultados apoyan las observaciones previas de que las nucleasas Cas12a son principalmente efectivas contra sitios que albergan PAMs TTTV (Kim et al., Nat Biotechnol., 2016, 34:863-8), y que no se ha demostrado que ningún ortólogo natural de Cas12a caracterizado hasta la fecha supere este requisito restrictivo de PAM en células humanas.

Para ampliar el rango de selección y la utilidad de las nucleasas Cas12a, aprovechamos los estudios estructurales del complejo ternario de AsCas12a (Yamano et al., Cell. 2016 May 5; 165(4):949-62) para diseñar una única variante capaz de reconocer PAMs canónicos y no canónicos. Se identificaron residuos en estrecha proximidad espacial a las bases de ADN de PAM (Fig. 20A), y se planteó la hipótesis de que la sustitución de estos residuos podría alterar

o relajar el reconocimiento de PAM mediante la creación de nuevas interacciones de base o no específicas. Primero probamos las actividades de variantes que codifican sustituciones únicas en estas posiciones en células humanas contra sitios que codifican PAMs canónicos y no canónicos. En comparación con AsCas12a de tipo salvaje, cuatro variantes de sustitución única (S170R, E174R, S542R, K548R) mostraron una actividad superior en los sitios PAM TTTA o TTTC canónicos, al tiempo que permitían una selección más eficiente de los sitios con PAMs CTTA o TTCC no canónicos (Figs. 20B y 20C, respectivamente).

Las pruebas combinatorias de estas sustituciones en células humanas revelaron aumentos sustanciales de la actividad en comparación con AsCas12a de tipo salvaje en cuatro sitios adicionales con PAMs no canónicos (ATTC, CTTA, GTTC y TTCC), y recapitulaban la observación de la mejora de la actividad en un sitio PAM TTTG canónico (Fig. 15A). Algunos de los aumentos más destacados de la actividad y las ampliaciones del rango de selección se observaron cuando se combinaron las sustituciones E174R y S542R, ya que las variantes E174R/S542R y E174R/S542R/K548R mostraron entre 4 y 32 veces más actividad en los sitios PAM no canónicos en comparación con AsCas12a, y casi 2 veces más actividad en el sitio PAM canónico (Fig. 20D). Por lo tanto, seleccionamos estas dos variantes para una mayor caracterización.

Para perfilar ampliamente las preferencias ampliadas de PAM de nuestras variantes de AsCas12a, optimizamos un ensayo de determinación imparcial de PAM de alto rendimiento *in vitro* (PAMDA; Figs. 21A-21H). Primero purificamos y ensayamos AsCas12a de tipo salvaje y E174R/S542R/K548R (Fig. 21A). Las actividades de escisión *in vitro* de estas nucleasas Cas12a se verificaron en sustratos plasmídicos que codificaban dos secuencias espaciadoras distintas con varios PAM seleccionables y no seleccionables, corroborando las mejoras de actividad observadas previamente en ensayos con células humanas (Figs. 21B y 21C). A continuación, adaptamos este flujo de trabajo para realizar el PAMDA mediante la construcción de dos bibliotecas de plásmidos separadas que codifican las mismas dos secuencias espaciadoras, pero que ahora albergan una secuencia aleatoria de 8-mer en lugar de PAM (Fig. 21D). Se llevaron a cabo reacciones de escisión en función del tiempo en las dos bibliotecas de plásmidos linealizados utilizando complejos de ribonucleoproteína (RNP) AsCas12a/ARNcr, seguidas de amplificación por PCR y secuenciación de los sustratos no escindidos para calcular las velocidades a las que se agotan los PAMs seleccionables (Fig. 21D). Se observaron fuertes correlaciones entre las constantes de velocidad específicas de los PAM (k ; para el agotamiento de los PAM de la población a lo largo del tiempo) en las secuencias PAM de 4 nt más próximas al espaciador entre réplicas y espaciadores a través de experimentos PAMDA separados (Figs. 21E y 21F, respectivamente). La agrupación del recuento de valores $\log_{10}k$ para cada uno de los posibles 256 PAMs de 4 nt tanto para AsCas12a E174R/S542R/K548R como de tipo salvaje sugirió un umbral aproximado para el reconocimiento y selección de PAM *bona fide* en el rango de $-2,25 \log_{10}k$ (Fig. 21G). El análisis de las curvas de agotamiento a partir de los datos PAMDA para las mismas combinaciones PAM/espaciador utilizadas para la optimización del ensayo *in vitro* utilizando sustratos PAM estáticos reveló perfiles de escisión consistentes (Figs. 21H).

Para realizar el PAMDA, las nucleasas Cas12a purificadas se combinan con ARNcr para interrogar a las bibliotecas de plásmidos que presentan secuencias aleatorias de 8 nt en lugar de PAM, lo que permite el cálculo de las constantes de velocidad *in vitro* (k) para el agotamiento de los PAMs seleccionables de la población. Los gráficos de los valores $\log_{10}k$ medios para AsCas12a de tipo salvaje en todas las posibles secuencias PAM de 4 nt revelaron que, como era de esperar, la selección sólo era eficiente en sitios con PAMs TTTV (Fig. 15B). Por el contrario, la variante E174R/S542R/K548R mostró una selección drásticamente ampliada de las clases PAM que incluían TTTN y TTCN (TTYN); ATTV, CTTV y GTTV (VTTV); TATV y TGTV (TRTV); y muchos PAMs adicionales (Fig. 15B). Como factor importante, este análisis también respaldó nuestra observación de que la variante mantiene un reconocimiento potente de PAMs TTTV canónicos.

A continuación, para obtener una comprensión más completa de las mejoras del rango de selección conferidas por cada sustitución, desplegamos el PAMDA en las variantes intermedias de sustitución simple y doble necesarias para generar E174R/S542R/K548R (Fig. 22A). En consonancia con los datos de nuestro ensayo en células humanas (Fig. 15A), este análisis reveló que la variante E174R/S542R también mostraba mejores actividades en una amplia gama de PAMs. Una comparación de los valores medios de PAMDA $\log_{10}k$ para E174R/S542R y E174R/S542R/K548R en PAMs NNYN demostró que ambas variantes poseen rangos de selección ampliados (Fig. 22B), lo que sugiere que las sustituciones E174R y S542R son responsables de gran parte del rango de selección ampliado. Curiosamente, las identidades de estos residuos no son compartidas por todos los ortólogos de Cas12a, sino que existen en regiones donde los aminoácidos que los flanquean están estrictamente conservados (Fig. 22C).

Para explorar un poco más las mejoras en el rango de selección que presentan las variantes E174R/S542R y E174R/S542R/K548R en células humanas, caracterizamos sus actividades en sitios que PAMDA identificó como seleccionables o no seleccionables, incluyendo 75 sitios VTTN y TTCN que albergaban PAMs que ahora deberían ser accesibles mayoritariamente con cualquiera de las variantes (VTTT como controles negativos; Fig. 23A), y 17 sitios con PAMs TATN donde los sitios TATV deberían ser seleccionados efectivamente sólo con

E174R/S542R/K548R (Fig. 23B). Como predecían los resultados de PAMDA para las variantes, observamos una selección consistente y robusta con E174R/S542R y E174R/S542R/K548R en sitios con PAMs VTTV y TTCN, una modificación ineficaz de los sitios VTTT, y sólo una selección eficaz de los sitios TATV con la variante E174R/S542R/K548R (Fig. 15C, 23A y 23B). Como factor importante, ambas variantes fueron mucho más efectivas en la selección de estos sitios PAM no canónicos en comparación con AsCas12a de tipo salvaje (Figs. 15D, 19A, y 23C). Dado que los resultados de PAMDA para la variante E174R/S542R/K548R indicaban que también podría reconocer potencialmente un rango más amplio de PAMs más allá de los que ya habíamos probado, examinamos 15 sitios que albergaban PAMs TGTV y otros 83 sitios en células humanas con PAMs alternativos en o cerca de un umbral de PAMDA $\log_{10}(k)$ medio de -2,25 (Figs. 23D y 23E, respectivamente). Observamos una modificación robusta de muchos de los sitios que albergaban estos PAMs no canónicos adicionales, y también una fuerte correlación entre las actividades medias de células humanas y los ks de PAMDA para la mayoría de las PAMs (Fig. 23F).

Una observación adicional del PAMDA fue que las variantes E174R/S542R y E174R/S542R/K548R podían ahora seleccionar PAMs TTTT antes inaccesibles con AsCas12a de tipo salvaje (Fig. 22A). Para determinar si estas variantes podían seleccionar eficazmente sitios con PAMs TTTT no canónicos, manteniendo al mismo tiempo su actividad en PAMs TTTV canónicos, comparamos sus actividades en 25 sitios TTTN adicionales en células humanas (Fig. 23G). En consonancia con nuestros hallazgos anteriores (Fig. 15A), observamos un aumento de aproximadamente 2 veces en la modificación de los sitios que portaban cada PAM TTTT, así como una gran mejora en la selección de los sitios que codifican PAMs TTTT (Fig. 15E). Estos resultados sugieren que las variantes con la combinación de E174R y S542R no sólo mejoran drásticamente el rango de selección, sino que también pueden mejorar sorprendentemente la selección de sitios con PAMs TTTN (Fig. 23H).

En general, la variante E174R/S542R/K548R, en lo sucesivo denominada AsCas12a mejorado (eAsCas12a), permite una expansión espectacular del rango de selección y una mejora de la actividad *on-target*. Los PAMs ahora accesibles con eAsCas12a pueden clasificarse en niveles de confianza basados en la consistencia entre los datos de PAMDA y de células humanas (Figs. 15F y 23I). Observamos una fuerte correlación entre el porcentaje medio de modificación en células humanas y el PAMDA ks medio determinado *in vitro* (Fig. 23F), lo que sugiere que el PAMDA es razonablemente predictivo de los PAMs seleccionables y no seleccionables en células humanas. Cabe señalar que los datos de PAMDA se generaron a partir de bibliotecas que codificaban dos secuencias espaciadoras distintas, por lo que es posible que los perfiles de preferencia de PAM observados en estas bibliotecas no representen las clasificaciones de PAM en todas las secuencias espaciadoras (aunque observamos una buena correlación entre las dos bibliotecas espaciadoras examinadas; Fig. 21F).

Así pues, clasificamos los PAMs que alcanzan un umbral estricto de más del 20% de selección media en células humanas en todos los sitios examinados y un PAMDA k superior a 0,01 como PAM de "nivel 1" (TTYN, CTTV, RTTC, TATM, CTCC, TCCC y TACA), y PAMs que alcanzan un umbral de selección media superior al 10% en células y un PAMDA k superior a 0,005 como PAMs de "nivel 2" (RTTS, TATA, TGTV, ANCC, CVCC, TGCC, GTCC, TTAC) (Fig. 23I). Los PAMs discrepantes (escasa correlación entre los datos de células humanas y los datos de PAMDA) y aquellos con una modificación media en células humanas inferior al 10% se clasifican como PAMs de "nivel 3" y no se recomiendan para la mayoría de las aplicaciones de edición genómica dados nuestros datos actuales. Para aplicaciones en las que el rango de selección puede no ser limitante y la eficiencia es el objetivo principal, recomendamos priorizar los PAMs dentro del nivel 1 o 2 basado en sus clasificaciones de PAMDA y células humanas (Fig. 23I).

En conjunto, eAsCas12a y otras variantes de AsCas12a mejoran la selección en más de 8 veces, permitiendo una selección de mayor resolución de regiones codificantes y no codificantes del genoma (Fig. 15G).

Actividad *on-target* mejorada con eAsCas12a

Más allá del rango de selección, otra propiedad crítica de las nucleasas de edición genómica es la potente actividad *on-target*. Por consiguiente, tratamos de entender mejor qué sustituciones contribuyen a nuestras observaciones de una mayor eficacia de selección con eAsCas12a, ya que hasta donde sabemos, no se han descrito sustituciones de aminoácidos que aumenten la eficacia de edición de las nucleasas CRISPR. Así, primero determinamos si eAsCas12a o sus variantes derivadas podían revertir las deficiencias de escisión del ADN a temperaturas más bajas descritas previamente para AsCas12a de tipo salvaje (Moreno-Mateos et al., Nat Commun., 2017, 8:2024). Las reacciones comparativas de escisión *in vitro* a 37, 32 y 25 °C revelaron que eAsCas12a casi elimina las diferencias de escisión dependientes de la temperatura observadas entre AsCas12a y LbCas12a, y que la recuperación fenotípica es atribuible en gran medida a las sustituciones E174R y S542R (Fig. 16A).

Una propiedad única de las nucleasas Cas12a es su capacidad para procesar ARNcr individual a partir de transcripciones de poli-ARNcr (Fonfara et al., Nature, 2016, 532:517-21), lo que simplifica la selección multiplex en las células (Zetsche et al., Nat Biotechnol., 2017, 35:31-34; Tak et al., Nat Methods, 2017, 14:1163-1166). Para

evaluar si las actividades mejoradas de eAsCas12a podían extenderse a la selección multiplex, comparamos las actividades de As, eAs y LbCas12a cuando se programaban con matrices de poli-ARNcr que codificaban cada uno tres ARNcr dirigidos a genes separados en células humanas (Figs. 18A-18C). En la mayoría de los casos, observamos una mejor selección con eAsCas12a cuando las matrices de ARNcr se expresaban a partir de un promotor de ARN polimerasa III, presumiblemente debido a la mayor actividad de eAsCas12a en sitios con PAMs canónicos (Fig. 18A y 18B). Esta mejora de la selección multiplex también se observó cuando el ARNcr se expresó a partir de un promotor de ARN polimerasa II, ampliando el alcance de las aplicaciones de edición multiplex (Fig. 18C). También diseñamos matrices multiplex que codificaban dos conjuntos de ARNcr seleccionados proximalmente para generar pequeñas deleciones genómicas. Los pares de ARNcr se expresaron a partir de transcripciones de poli-ARNcr o transfectando grupos de plásmidos de ARNcr individuales en células, y observamos de nuevo una mejor selección multiplex con eAsCas12a (Fig. 18D).

El ejemplo 4 proporciona pruebas adicionales que apoyan la observación de que la sustitución E174R mejora la actividad *on-target*.

Ejemplo 2. Variantes de LbCpfI con especificidad PAM alterada

Dado que AsCpfI y LbCpfI comparten un alto grado de homología a través de los residuos en la vecindad de los contactos proteína-ADN que rodean al PAM (en base a estructuras cristalinas tridimensionales y un alineamiento primario de secuencias, ver la Tabla 1), creamos variantes de PAM LbCpfI que albergarían residuos en posiciones que corresponderían a las posiciones a las que mutamos en AsCpfI. Las sustituciones simples en las posiciones T152, D156, G532 y K538 en LbCpfI (que corresponden a los residuos S170, E174, S542 y K548 en AsCpfI) revelaron sólo aumentos modestos en la actividad de disrupción de EGFP frente a sitios con PAMs no canónicos (Fig. 12A) en comparación con sus correspondientes variantes AsCpfI (Fig. 5). Sin embargo, la variante de triple sustitución LbCpfI-D156R/G532R/K538R (análoga a AsCpfI-E174R/S542R/K548R) mostró un aumento ligeramente más sustancial en la selección de sitios con PAMs no canónicos en comparación con LbCpfI de tipo salvaje en el ensayo de disrupción de EGFP (Fig. 12B).

Ejemplo 3. Variantes de FnCpfI con especificidad PAM alterada

Informes anteriores han sugerido que FnCpfI no funciona, o tiene poca actividad en células humanas (Zetsche, Cell 2015; Kim, Nature Biotechnology 2016). Dado que AsCpfI y FnCpfI comparten un alto grado de homología, primero tratamos de probar si FnCpfI de tipo salvaje funciona realmente en células humanas y, después, si podíamos relajar la especificidad PAM de TTN notificada previamente (Zetsche, Cell 2015).

Para examinar la actividad de FnCpfI en células humanas, probamos su actividad en nuestro ensayo de disrupción de EGFP en células humanas contra sitios diana que contenían PAMs de la forma NTTN, TNTN y TTTN (Fig. 13A). Nuestros resultados revelan que el FnCpfI de tipo salvaje puede efectivamente mediar una disrupción robusta de EGFP en células humanas contra sitios NTTN, con cierta actividad detectable contra sitios TCTN y TTCN (Fig. 13A). A continuación, comparamos la actividad de disrupción génica endógena de FnCpfI en AsCpfI y LbCpfI en 10 sitios diana endógenos diferentes con PAMs TTTN. En muchos casos, observamos una actividad de FnCpfI comparable a la de AsCpfI y LbCpfI, lo que demuestra que FnCpfI funciona realmente de forma robusta en células humanas (Fig. 13B).

Dado que FnCpfI funciona en células humanas, intentamos determinar si podíamos generar variantes de PAM FnCpfI creando variantes de FnCpfI que portaran sustituciones en residuos homólogos en posiciones de AsCpfI que condujeran a una especificidad de PAM alterada (Tabla 1). De las sustituciones que examinamos, las sustituciones únicas de K180R, N607R y D616R condujeron a aumentos de la actividad sobre AsCpfI de tipo salvaje en los sitios PAM TTTN, TNTN y NTTN (Fig. 13C). Además, una mutación K671H podría aumentar la actividad contra un sitio PAM TCTN. También observamos que las variantes que portaban combinaciones de sustituciones que incluían N607R/K613R, N607R/K613V, N607R/K613V/D616R, o N607R/K613R/D616R mejoraban la actividad sobre el FnCpfI de tipo salvaje en ciertos PAMs de la forma TTTN, CTTN, GTTN, TATN, TCTN, TCTN, TTAN, de TTCN (Fig. 13C).

Ejemplo 4. Variantes adicionales de AsCpfI con especificidades PAM alteradas

Gao et al. publicaron recientemente otras variantes de CpfI con especificidad PAM alterada (Gao et al., "Engineered CpfI Enzymes with Altered PAM Specificities", bioRxiv 091611; doi: <https://doi.org/10.1101/091611>). Estas variantes, con sus actividades reivindicadas en PAMs canónicos y/o no canónicos, son las siguientes:

- 1) AsCpfI-S542R/K548V/N552R - funciona contra sitios PAM TATV

2) AsCpf1-S542R/K607R - funciona de forma óptima contra sitios PAM TYCV, pero muestra una pérdida de actividad contra sitios TTTV canónicos.

Para contrastar las variantes de triple y cuádruple sustitución de AsCpf1 aquí descritas (E174R/S542R/K548R y E174R/S542R/K548R/N551R, respectivamente) con la variante S542R/K548V/N552R, comparamos la actividad de estas tres variantes utilizando el ensayo de interrupción EGFP en sitios diana que presentaban TTTN canónico, TATN (que según los informes reconoce la variante S542R/K548V/N552R) y PAMs con diferencias de una o dos bases (Fig. 14A). En todos los sitios probados, observamos que nuestras variantes de sustitución triple y cuádruple superaban a la variante S542R/K548V/N552R en TTTV, TATN y otros PAMs no canónicos CTTN, GTTN, TCTN, TGTN, TTAN, TTCN, TTGN y TCCN (Fig. 14A). A continuación, basándonos en nuestras observaciones previas de que las sustituciones S170R o E174R pueden aumentar la actividad de las variantes de AsCpf1 cuando se combinan con otras sustituciones, exploramos si la adición de cualquiera de estas sustituciones a la variante S542R/K548V/N552R también podría mejorar su actividad. Al comparar las variantes de cuádruple sustitución S170R/S542R/K548V/N552R y E174R/S542R/K548V/N552R con S542R/K548V/N552R parental, observamos que la adición de las sustituciones S170R o E174R mejoraba sustancialmente la actividad (siendo el efecto de E174R mayor que el de S170R, aunque la adición de S170R también produce mejoras; Fig. 14A).

También comparamos la actividad de nuestras variantes AsCpf1 de sustitución triple y cuádruple (E174R/S542R/K548R y E174R/S542R/K548R/N551R, respectivamente) frente a la variante S542R/K607R en una serie de sitios diana en EGFP con la secuencia PAM TTTN canónica o PAMs con una sola base de diferencia. Para todos estos sitios, nuestras variantes de sustitución triple y cuádruple (E174R/S542R/K548R y E174R/S542R/K548R/N551R) tenían niveles aproximadamente iguales o superiores de actividad de interrupción de EGFP en comparación con la variante S542R/K607R (Fig. 14B). S542R/K607R sólo superó a nuestras variantes triple y cuádruple en un sitio diana con un PAM TCCN. Por tanto, añadimos las sustituciones S170R o E174R a la variante S542R/K607R para crear las variantes de triple sustitución S170R/S542R/K607R y E174R/S542R/K607R. Estas variantes adicionales de triple sustitución funcionaron tan bien o mejor que la variante S542R/K607R en todos los sitios probados, especialmente en los sitios con un PAM TCCN (Fig. 14B). Por último, comparamos una de nuestras variantes de triple sustitución (E174R/S542R/K548R) con la variante S542R/K607R en una serie de sitios diana de genes humanos endógenos en células U2OS con PAMs no canónicos (con 1, 2 o 3 diferencias de base en el PAM). En todos los sitios probados distintos de los que portaban PAMs NCCC, nuestra variante E174R/S542R/K548R funcionó tan bien o mejor que la variante S542R/K607R (Fig. 14C). Para los sitios PAM NCCC en los que la variante S542R/K607R mostraba una mayor actividad de interrupción génica, comparamos esta variante con nuestras variantes S170R/S542R/K607R, E174R/S542R/K607R y E174R/S542R/K607H y observamos que en todos los casos, una de nuestras variantes de triple sustitución superaba a la variante S542R/K607H (Fig. 14C).

Ejemplo 4B. Mejora de las actividades *on-target* de las variantes PAM de AsCas12a

Dado que nuestros resultados sugieren que E174R y S542R conducen a actividades mejoradas de eAsCas12a en células humanas, planteamos la hipótesis de que la inclusión de E174R en las variantes de AsCas12a previamente descritas que ya codifican S542R también podría mejorar sus actividades. Así, la sustitución E174R se combinó con las variantes RVR (S542R/K548V/N552R) y RR (S542R/K607R) para crear versiones mejoradas de estas nucleasas (eRVR y eRR, respectivamente). La comparación de las actividades de las variantes eAsCas12a, RVR, eRVR, RR y eRR contra 11 sitios con PAMs TTTN en células humanas (Fig. 24A) reveló que, mientras que las variantes RVR y RR publicadas anteriormente tienen actividades similares o más débiles en comparación con el AsCas12a de tipo salvaje, la adición de E174R para crear las variantes eRVR y eRR dio lugar a aumentos de más de 2 veces en sus actividades (aunque aún más bajas que eAsCas12a; Fig. 16B). Estos resultados refuerzan la observación de que las variantes con la combinación de E174R y S542R pueden mejorar la actividad *on-target*.

A continuación, debido a que nuestra evaluación PAMDA de eAsCas12a reveló el reconocimiento de PAMs primarios previamente notificados como accesibles por las variantes RVR y RR (PAMs TATV y TYCV, respectivamente; Fig. 15B), comparamos eAsCas12a con las versiones publicadas y mejoradas de estas nucleasas en dichos sitios en células humanas. En 12 sitios TATN (Fig. 24B), observamos que eAsCas12a mostraba una actividad aproximadamente equivalente a la variante RVR (Fig. 16C). Curiosamente, la adición de E174R a RVR condujo a una mejora de 2 veces en la actividad, lo que sugiere eRVR como la variante óptima para aplicaciones en las que la selección de los sitios TATN es el objetivo principal (Fig. 16C). Después evaluamos eAsCas12a, RR y eRR en 29 sitios con PAMs TYCN en células humanas (Fig. 24C). eAsCas12a mostró una mayor modificación en comparación con RR en los 18 sitios TTCN, mientras que la variante eRR que contenía E174R tuvo una actividad comparable a eAsCas12a (Fig. 16D). Una comparación más detallada de estas variantes en 11 sitios TCCN reveló que, si bien la variante RR es una nucleasa más eficaz en comparación con eAsCas12a en sitios con PAMs TCCN, una vez más la variante eRR que contiene E174R dio lugar a la modificación más robusta en todos los sitios TCCN (Fig. 16D).

Para determinar si el rango de selección de las variantes mejoradas eRVR y eRR se había visto alterado por la adición de E174R, aplicamos el PAMDA a RVR, RR y sus variantes mejoradas (Fig. 24D). En consonancia con nuestros datos en células humanas, observamos que las nucleasas eRVR y eRR tenían un rango de selección similar al de sus variantes RVR y RR parentales, pero que su potencia *on-target* mejoraba con la adición de la sustitución E174R (Fig. 16E y 24E). En conjunto, estos resultados demuestran que las sustituciones E174R y S542R no sólo mejoran el rango de selección, sino que también pueden mejorar las actividades *on-target* de las nucleasas AsCas12a.

Ejemplo 5. Aumento de la especificidad genómica de eAsCas12a

Dado que eAsCas12a presenta una actividad mejorada y un reconocimiento relajado de PAM en comparación con AsCas12a de tipo salvaje, buscamos comparar las especificidades de estas nucleasas ya que su capacidad para distinguir los sitios *on-* y *off-target* es crítica tanto para la investigación como para las aplicaciones terapéuticas. A este respecto, nosotros y otros hemos demostrado previamente que las nucleasas Cas12a de tipo salvaje poseen especificidades robustas en todo el genoma y son relativamente intolerantes a los sitios *off-target* discordantes que albergan discordancias erróneas simples o dobles en las regiones inmediatamente PAM proximal, media y PAM distal del espaciador (Kleinstiver et al., Nat Biotechnol., 2016, 34:869-74; Kim et al., Nat Biotechnol., 2016, 34:863-8; WO2018/022634). Por tanto, utilizamos el método de identificación no sesgada de DSBs en todo el genoma mediante secuenciación (GUIDE-seq) (Tsai et al., Nat Biotechnol., 2015, 33:187-197) para comparar las especificidades en todo el genoma de As y eAsCas12a en cuatro sitios con PAMs TTTV (Figs. 25A-25D). GUIDE-seq detectó pocos *off-targets* con AsCas12a de tipo salvaje, y observamos un aumento en el número de *off-targets* para eAsCas12a (Figs. 17A y 17B). Muchos de los *off-targets* observados para eAsCas12a habían sido identificados previamente en experimentos GUIDE-seq con LbCas12a (Kleinstiver et al., Nat Biotechnol., 2016, 34:869-74), contenían discordancias en posiciones que se sabe que toleran sustituciones de nucleótidos (Kleinstiver et al., Nat Biotechnol., 2016, 34:869-74; Kim et al., Nat Biotechnol., 2016, 34:863-8), o codifican PAMs no canónicos ahora seleccionables (Fig. 17A).

Para explorar si un método de evolución dirigida podría utilizarse para mejorar la fidelidad de eAsCas12a, examinamos el impacto de las sustituciones de aminoácidos en residuos de AsCas12a que se prevé que hagan contactos no específicos con el ADN. Evaluamos la tolerancia a una sola discordancia de nueve sustituciones de alta fidelidad (HF) putativas (incluyendo la variante K949A descrita previamente), y observamos que mientras algunas sustituciones mejoraban el perfil de tolerancia a una sola discordancia de AsCas12a en dos sitios, muchas también reducían la actividad con el ARNcr coincidente (Fig. 25E). Combinamos las sustituciones más prometedoras con eAsCas12a, y observamos que la versión N282A de eAsCas12a (denominada eAsCas12a-HF1) producía las mejoras más deseables en la intolerancia a la discordancia única y el mantenimiento de la actividad *on-target* (Fig. 25F). La evaluación de eAsCas12a y eAsCas12a-HF1 utilizando el PAMDA reveló perfiles de preferencia de PAM casi idénticos (Figs. 25G y 25H), lo que sugiere que la mutación N282A HF no altera el reconocimiento de PAM o el rango de selección (Fig. 17C).

A continuación, para determinar si eAsCas12a-HF1 puede mejorar la especificidad en todo el genoma, realizamos GUIDE-seq utilizando los mismos cuatro ARNcr diana de PAM TTTV previamente examinados. Comparado con eAsCas12a, observamos una reducción tanto en el número como en la frecuencia con la que se detectaban *off-targets* con eAsCas12a-HF1 para 3 de los 4 ARNcr (Figs. 17A y 17B), donde sus perfiles de especificidad ahora se parecían más a los observados para AsCas12a de tipo salvaje. Se realizaron experimentos GUIDE-seq adicionales para comparar eAsCas12a y eAsCas12a-HF1 en sitios con PAMs no canónicos (Figs. 17D y 25B-25D), y de nuevo observamos reducciones en el número y la frecuencia de sitios *off-targets* con eAsCas12a-HF1 en comparación con eAsCas12a (Figs. 17E y 17F, respectivamente).

Comparamos entonces las actividades *on-target* de eAsCas12a y eAsCas12a-HF1 en sitios PAM canónicos y no canónicos (Figs. 25I y 25J, respectivamente) para examinar si la sustitución N282A afecta a la eficacia de la selección. Observamos una modificación génica similar en 8 sitios PAM TTTN (de nuevo con una eficiencia casi 3 veces mayor en relación con AsCas12a de tipo salvaje; Fig. 17G), y actividades comparables entre eAsCas12a y eAsCas12a-HF1 en 15 sitios con PAMs no canónicos (Fig. 17H). Además, los ensayos de escisión *in vitro* para evaluar la tolerancia a la temperatura revelaron perfiles de escisión similares entre eAsCas12a, eAsCas12a-HF1 y LbCas12a a 37, 32 y 25 °C (Fig. 25K). En conjunto, estos resultados demuestran que eAsCas12a-HF1 puede mejorar la especificidad en todo el genoma manteniendo al mismo tiempo un importante rango de selección, una mayor actividad y propiedades de tolerancia a la temperatura.

Ejemplo 6. Aprovechamiento de las propiedades mejoradas de eAsCas12a para la activación de genes y aplicaciones de edición del epigenoma

Otra adaptación destacada de CRISPR-Cas12a ha sido la edición del epigenoma, donde se ha demostrado que las fusiones de Cas12a (dCas12a) inactiva DNasa a efectores heterólogos modulan la expresión génica. Anteriormente

demostramos que las fusiones dLbCas12a al dominio sintético de transactivación VPR (una combinación de VP64, p65 y Rta; Chavez et al., *Nat Methods.*, 2015, 12:326-8) mediaban una activación génica más potente comparado con fusiones dAsCas12a equivalentes en células humanas (Tak et al., *Nat Methods*, 2017, 14:1163-1166). Para explorar si eAsCas12a podría mejorar la edición del epigenoma en comparación con los efectores LbCas12a, primero diseñamos y probamos diferentes configuraciones de fusiones dAs, deAs y dLbCas12a a VPR (Fig. 26A). Las comparaciones de sus actividades en los sitios PAM TTTV canónicos y TTCV no canónicos proximales al promotor de VEGFA establecieron una arquitectura de fusión óptima dCas12a-VPR (versión 1.1; Figs. 26B-26D, y también revelaron que los efectores deAs-VPR facilitaban una mayor producción de VEGFA en relación con las fusiones dAs y dLbCas12a cuando se utilizaban ARNcr dirigidos a sitios canónicos o no canónicos (Fig. 26E). En experimentos en los que se comparó deAs-VPR con la fusión prototípica dSpCas9-VPR (dirigida a sitios separados pero cercanos), observamos una activación génica comparable o mejor con deAs-VPR (Figs. 26B-C y 26E-F).

Experimentos adicionales con fusiones dAs, deAs y dLb-VPR dirigidas a sitios en los promotores de tres genes endógenos adicionales (NPY1R, HBB y AR) revelaron una vez más la activación génica más potente con deAs-VPR cuando se utilizaban grupos de PAM canónicos (Fig. 18E) o PAM no canónicos dirigidos a ARNcr (Fig. 18F y 18G). La fusión deAs-VPR logró entre 10 y 10.000 veces la activación génica, superando con frecuencia a dAs o dLbCas12a-VPR en más de 10-100 veces.

En conjunto, la fusión deAsCas12a a VPR puede mediar en la activación robusta de genes con eficiencias equivalentes o superiores en comparación con las fusiones dLbCas12a-VPR publicadas cuando se dirige a sitios PAM TTTV canónicos, y también ofrece la capacidad novedosa de activar genes por selección de sitios PAM no canónicos accesibles sólo con esta variante eAsCas12a. Estos resultados recapitulan las propiedades de mayor actividad y mejor rango de selección de eAsCas12a, y proporcionan tecnologías de activación génica potentes y de amplio rango de selección que también pueden adaptarse a otras aplicaciones de edición del epigenoma.

Ejemplo 7. Variantes de AsCas12a y LbCas12a para aplicaciones de edición de bases

Se ha demostrado recientemente la capacidad de realizar eventos de edición precisos de una sola base utilizando construcciones de editor de bases (BE) SpCas9 (ver, por ej., Komor et al., *Nature*. 2016 May 19;533(7603):420-4; Nishida et al., *Science*. 2016 Sep 16;353(6305); Kim et al., *Nat Biotechnol*. 2017 Abr;35(4):371-376; Dong et al., *Sci*. 2017 Aug 30;3(8):eaao4774; and Gaudelli et al., *Nature*. 2017 Nov 23;551(7681):464-471), que explotan la formación de bucles R formados por SpCas9-ARNg que provocan la accesibilidad del ADNss de la cadena de ADN no diana. La fusión de dominios enzimáticos heterólogos de citidina o adenina desaminasa a SpCas9 puede, por tanto, actuar sobre la cadena de ADNss expuesta, dando lugar a la introducción eficiente de C a T, o de A desaminase G, respectivamente. Dado que la reparación celular de escisión de bases (BER) emplea uracilo ADN glicosilasa (UDG; también conocido como uracilo N-glicosilasa, o UNG) para extirpar bases de uracilo, este proceso endógeno puede revertir eficazmente las ediciones generadas por BEs de citidina porque la desaminación de la citidina conduce a un intermediario de uracilo. Por tanto, para mejorar la eficacia de las BE de citidina, también se pueden fusionar dominios efectores heterólogos como el inhibidor de uracilo glicosilasa (UGI) a SpCas9 para inhibir UDG, subvirtiendo el inicio de la BER y aumentando la efectividad de BEs de citidina.

Dado que nuestras observaciones previas sugerían que eAsCas12a (E174R/S542R/K548R) posee una actividad mejorada, nos preguntamos si eAsCas12a podría permitir el desarrollo de editores de bases de AsCas12a putativos (AsBEs). Para probar esta hipótesis, clonamos cuatro arquitecturas diferentes de eAsBE inactivas a DNasa (BE-1.1-1.4; Fig. 18H) que incluían una fusión N-terminal de rAPOBEC1, una sustitución D908A para anular la actividad de hidrólisis del ADN mediada por nucleasa, y una fusión C-terminal de UGI, y comparamos sus actividades con AsBE1.1 y 1.4 de tipo salvaje utilizando ocho ARNcr diferentes. Observamos una edición mínima (<1%) de C-a-T con AsBEs a través de todos los Cs para 7 de los 8 sitios (Fig. 18I). Curiosamente, las fusiones eAsBE demostraron niveles absolutos mucho mayores de conversión de C-a-T en los mismos ocho sitios (rango de edición 2-34%; Figs. 18I y 18J), mejorando drásticamente la edición en relación con AsBE (Fig. 27A). La evaluación de dos arquitecturas LbBE equivalentes (rango de edición 2-19% C-a-T) reveló niveles comparables de edición C-a-T en relación con eAsBEs (Fig. 18I y 18J). Para todas las construcciones, las eficiencias de edición variaron según el sitio diana y la arquitectura BE (Fig. 18I), y de forma similar a lo observado con SpCas9BEs, la presencia de una G 5' de una C pareció amortiguar la edición C-a-T (Fig. 27B). Se observaron purezas de edición deseables (predominantemente de C-a-T como producto principal) con Cas12a-BEs para posiciones editadas a altas eficiencias (Fig. 27C). Se observaron bajos niveles de indels para Cas12a-BEs presumiblemente debido a la inactivación de su actividad DNasa (Fig. 18K y Fig. 27D). En conjunto, estos resultados demuestran que las actividades mejoradas de eAsCas12a permiten la edición C-a-T a niveles antes inalcanzables con AsBEs y con una eficacia comparable a LbBEs, y amplían el potencial de los reactivos CRISPR de edición de bases.

Ejemplo 8. Variantes de AsCas12a para la detección de ADN

Una aplicación adicional descrita recientemente de las nucleasas CRISPR-Cas12a se basa en la observación de que las moléculas Cas12a presentan actividad DNasa inespecífica programada por la diana (Chen et al., Science, 2018, doi: 10.1126/science.aar6245), una propiedad que se ha aprovechado para la detección sensible de moléculas específicas de ADN en solución (Chen et al., Science, 2018, doi: 10.1126/science.aar6245; Gootenberg et al., Science, 2018, doi: 10.1126/science.aag0179). Cuando el complejo Cas12a-ARNcr se une a un ADN diana, el sitio activo catalítico de la RuvC DNasa adopta una conformación hiperactiva que digiere indiscriminadamente el ADN cercano. Se puede añadir a la solución una molécula sintética de fluoróforo reporter de ADN atenuado, lo que facilita la cuantificación de la actividad Cas12a-DNasa que libera el reporter fluorescente (East-Seletsky, Nature, 2016, 538:270-273). Así pues, el rango de selección ampliado y las actividades mejoradas de eAsCas12a podrían mejorar potencialmente las metodologías de detección de ADN aumentando la sensibilidad, facilitando la detección de moléculas de ADN con PAMs no canónicos o permitiendo la detección de alelos variantes con fines de diagnóstico.

Por tanto, tratamos de comparar las actividades colaterales trans-DNasa de AsCas12a de tipo salvaje y eAsCas12a para evaluar la compatibilidad de nuestra variante modificada mediante ingeniería con los flujos de trabajo de detección de ADN. Ensamblamos complejos Cas12-ARNcr *in vitro* y los programamos con sustratos de ADN activadores (coincidentes con la molécula reporter) o no activadores (control) antes de añadir la molécula reporter. También variamos el PAM codificado en el sustrato de ADN activador para determinar si el rango de selección ampliado de eAsCas12a recapitulaba en este ensayo *in vitro*. En experimentos con AsCas12a de tipo salvaje, observamos una detección robusta en presencia del sustrato coincidente que codifica un sitio PAM TTTA canónico, y una actividad muy reducida cuando se programaba con un sustrato con un PAM ACCT no canónico (Fig. 28A). A continuación, tanto eAsCas12a como eAsCas12-HF1 mostraron niveles de detección comparables a los de AsCas12a de tipo salvaje en el sustrato PAM TTTA (Fig. 28B), pero también se pudo detectar de forma robusta un sustrato PAM CTTA no canónico (como era de esperar dado el perfil ampliado de preferencia de PAM de eAsCas12a (Fig. 15B). Sorprendentemente, la enzima eAsCas12a también fue capaz de detectar un sustrato de ADN con un PAM ACCT no seleccionable (Fig. 28B), lo que sugiere posibles diferencias en los requisitos de PAM para la escisión prototípica del ADN diana o para actividades trans-DNasa no específicas. Estos resultados demuestran que tanto eAsCas12a como eAsCas12a-HF1 son potentes nucleasas modificadas mediante ingeniería para la detección de ADN que ofrecen un rango de selección y potencialmente ventajas de especificidad sobre AsCas12a de tipo salvaje.

OTRAS REALIZACIONES

Debe entenderse que, si bien la invención se ha descrito junto con la descripción detallada de la misma, la descripción anterior pretende ilustrar y no limitar el alcance de la invención, que se define por el alcance de las reivindicaciones adjuntas. Otros aspectos, ventajas y modificaciones están dentro del alcance de las reivindicaciones siguientes.

REIVINDICACIONES

1. Una proteína CRISPR aislada de *Prevotella* y *Francisella* 1 (Cpfl) de *Acidaminococcus* sp. BV3L6 (AsCpfl), que comprende una secuencia que es al menos un 80% idéntica a la secuencia de aminoácidos de SEQ ID NO:2 con una mutación en la posición E174R de SEQ ID NO:2.
- 5 2. La proteína Cpfl aislada de la reivindicación 1, donde la secuencia es al menos un 95% idéntica a la secuencia de aminoácidos de SEQ ID NO:2.
3. La proteína Cpfl aislada de la reivindicación 1 o 2, que comprende además una mutación en S542, S170, K548, N551, T167, T539, N552, M604 y/o K607, comprendiendo preferentemente la mutación en S542Q, S542K, o S542R.
- 10 4. La proteína Cpfl aislada de la reivindicación 1 o la reivindicación 2 que comprende una o más de las siguientes mutaciones: E174R, S170R, K548R, S170K, E174K, T167A, T539K, T539R, K548V, N551R, N552R, M604A, K607Q, K607R, K607S, y/o K607H preferentemente, que comprende las siguientes mutaciones: E174R/K548R, E174R/S542R, E174R/S542R/K548R, E174R/S542R/K607R, E174R/S542R/K607H, E174R/S542R/K548R/N551R, E174R/S542R/K548V/N552R, o E174R/S542R/K607R de SEQ ID NO:2.
- 15 5. La proteína Cpfl aislada de cualquiera de las reivindicaciones 1-4, que comprende además una o más mutaciones que disminuyen la actividad de la nucleasa, preferentemente seleccionadas del grupo que consiste en mutaciones en D908, E993, R1226, D1235, y/o D1263, preferentemente D908A, E993A, y/o R1226A.
6. Una proteína de fusión que comprende la proteína aislada de las reivindicaciones 1-5, fusionada a un dominio funcional heterólogo, con un enlazador intermedio opcional, donde el enlazador no interfiere con la actividad de la proteína de fusión.
- 20 7. La proteína de fusión de la reivindicación 6, donde el dominio funcional heterólogo es un dominio de activación transcripcional; preferentemente donde el dominio de activación transcripcional es VP16, VP64, Rta, NF- κ B p65, o una fusión VPR; o donde el dominio funcional heterólogo es un silenciador transcripcional o un dominio de represión transcripcional, preferentemente,
- 25 donde el dominio de represión transcripcional es un dominio KRAB (*Krueppel-associated box*), un dominio represor ERF (ERD) o un dominio de interacción mSin3A (SID); o donde el silenciador transcripcional es la Proteína 1 Heterocromatina (HP1);
- o donde el dominio funcional heterólogo es una enzima que modifica el estado de metilación del ADN, preferiblemente donde la enzima que modifica el estado de metilación del ADN es un ADN metiltransferasa (DNMT) o una proteína TET, preferiblemente, donde la proteína TET es TET1;
- 30 o donde el dominio funcional heterólogo es una enzima que modifica una subunidad de histona, preferiblemente donde la enzima que modifica una subunidad de histona es una histona acetiltransferasa (HAT), histona desacetilasa (HDAC), histona metiltransferasa (HMT) o histona desmetilasa, preferiblemente donde el dominio funcional heterólogo es una biomolécula, preferiblemente donde la biomolécula es MS2, Csy4 o proteína lambda N; o, donde el dominio funcional heterólogo es FokI.
- 35 8. Un ácido nucleico aislado que codifica la proteína o proteína de fusión de las reivindicaciones 1-7.
9. Un vector que comprende el ácido nucleico aislado de la reivindicación 8.
10. El vector de la reivindicación 9, donde el ácido nucleico aislado codifica la proteína aislada de la reivindicación 1 y está enlazado operablemente a uno o más dominios reguladores para expresar una proteína CRISPR aislada de *Prevotella* y *Francisella* 1 (Cpfl) de *Acidaminococcus* sp. BV3L6 (AsCpfl), con mutaciones en una o más de las siguientes posiciones: S170, K548, N551, T167, T539, N552, M604, y/o K607 de SEQ ID NO:2.
- 40 11. Una célula huésped, preferentemente una célula huésped de mamífero, que comprende el ácido nucleico de la reivindicación 8, y opcionalmente expresa la proteína o proteína de fusión de las reivindicaciones 1-7.
12. Un método *ex vivo* o *in vitro* para alterar el genoma de una célula, el método comprende expresar en la célula, o contactar la célula con, la proteína aislada o proteína de fusión de la reivindicación 1-7, y al menos un ARN guía que tiene una región complementaria a una porción seleccionada del genoma de la célula, donde el método no es un método para modificar la identidad genética de la línea germinal de seres humanos.
- 45 13. La proteína aislada o proteína de fusión de las reivindicaciones 1-7 para el uso en un método terapéutico de alteración del genoma de una célula, el método que comprende expresar en la célula, o contactar la célula con,

dicha proteína aislada o proteína de fusión y al menos un ARN guía que tiene una región complementaria a una porción seleccionada del genoma de la célula, donde el método no es un método para modificar la identidad genética de la línea germinal de seres humanos.

- 5 14. La proteína aislada o proteína de fusión para el uso de la reivindicación 13, donde la proteína aislada o proteína de fusión comprende una o más de una secuencia de localización nuclear, secuencia peptídica de penetración celular, y/o marcador de afinidad;

o, donde la célula es una célula madre;
o donde la célula es una célula madre embrionaria, una célula madre mesenquimal o una célula madre pluripotente inducida;
10 o donde la célula se encuentra en un animal vivo;
o donde la célula se encuentra en un embrión no humano.
- 15 15. Un método *ex vivo* o *in vitro* para alterar una molécula de ADN de doble cadena (ADNs), el método comprende poner en contacto la molécula de ADNs con la proteína aislada o proteína de fusión de las reivindicaciones 1-7, y un ARN guía que tiene una región complementaria a una porción seleccionada de la molécula de ADNs, donde el método no es un método para modificar la identidad genética de la línea germinal de seres humanos.
- 20 16. La proteína aislada, o proteína de fusión de las reivindicaciones 1-7, para el uso en un método terapéutico de alteración de una molécula de ADN de doble cadena (ADNs), comprendiendo el método contactar la molécula de ADNs con dicha proteína aislada o proteína de fusión y un ARN guía que tiene una región complementaria a una porción seleccionada del ADN, donde el método no es un método para modificar la identidad genética de la línea germinal de seres humanos.

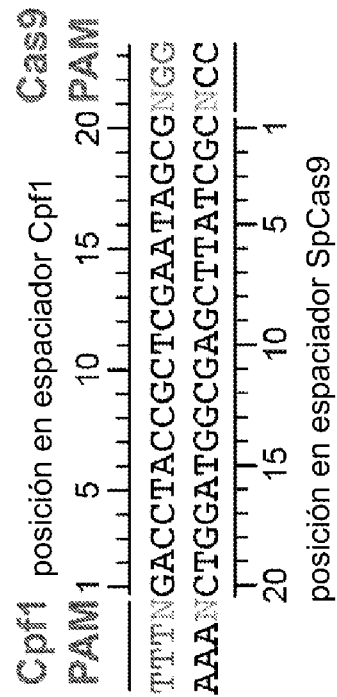


FIG. 1

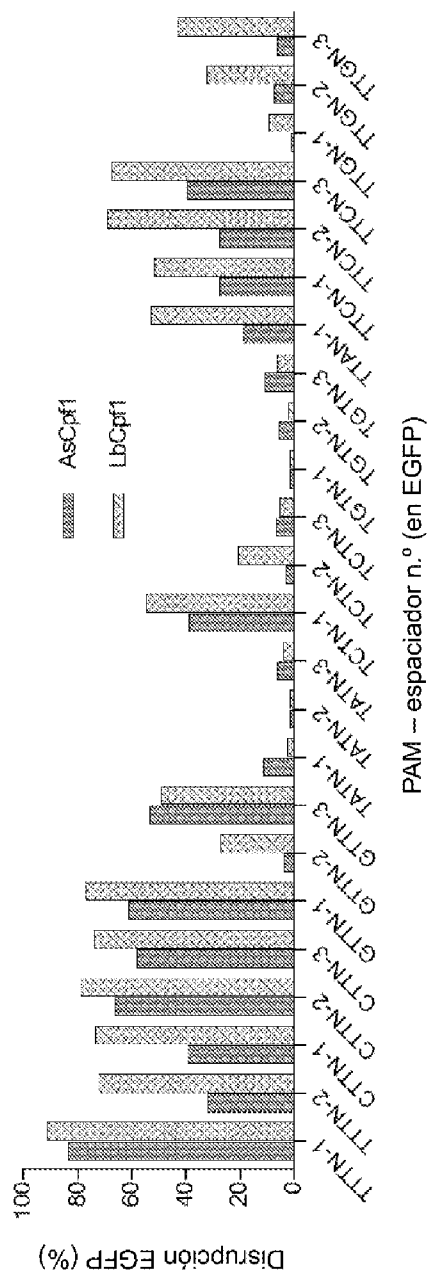


FIG. 2

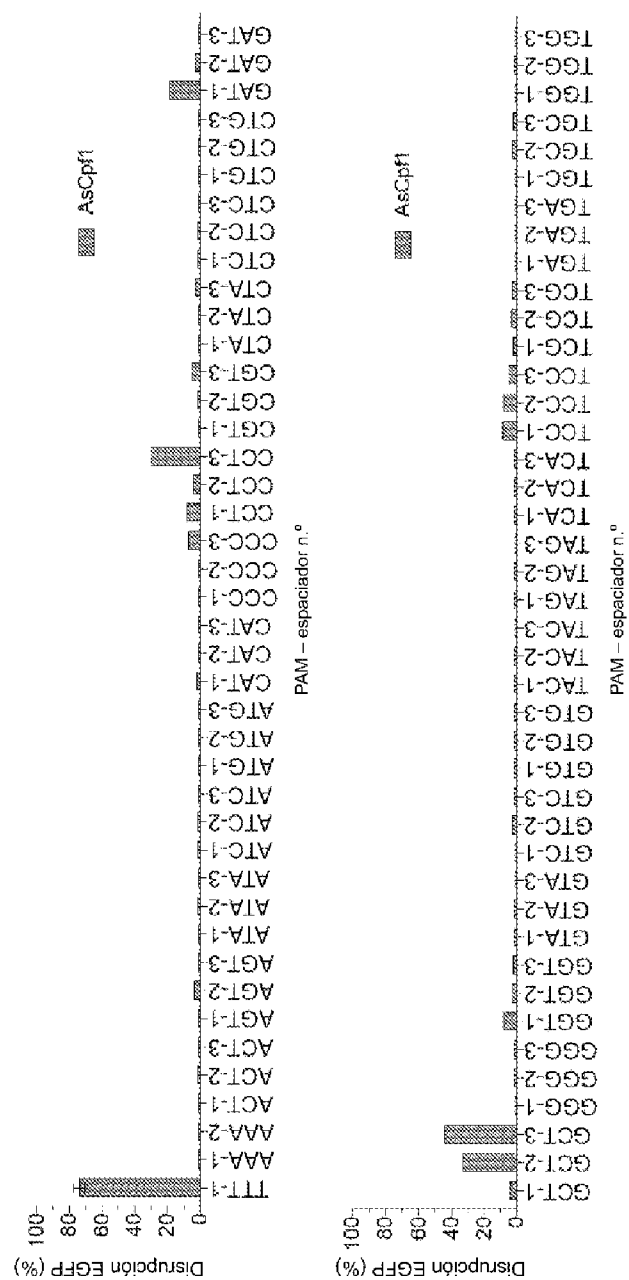
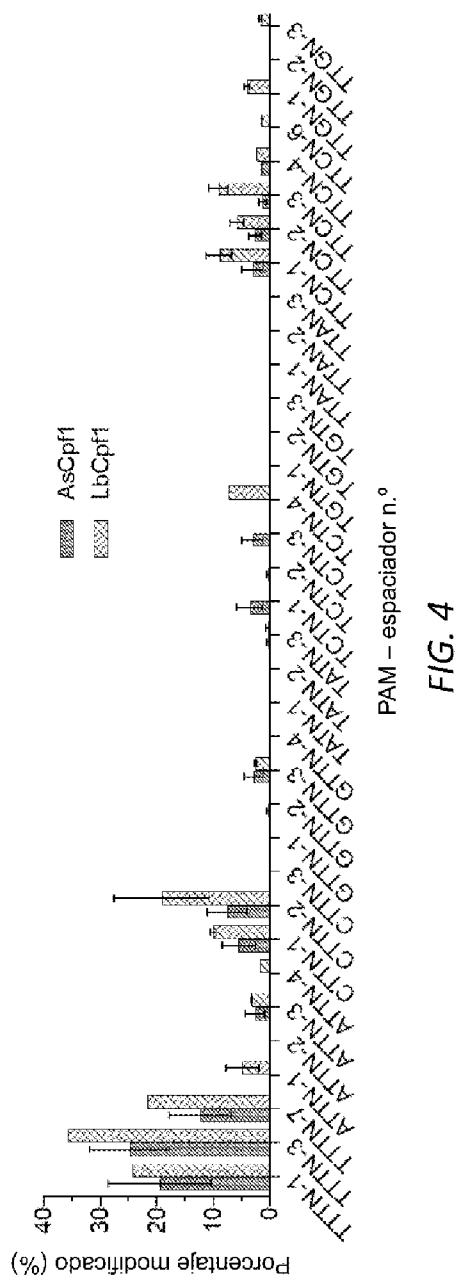


FIG. 3



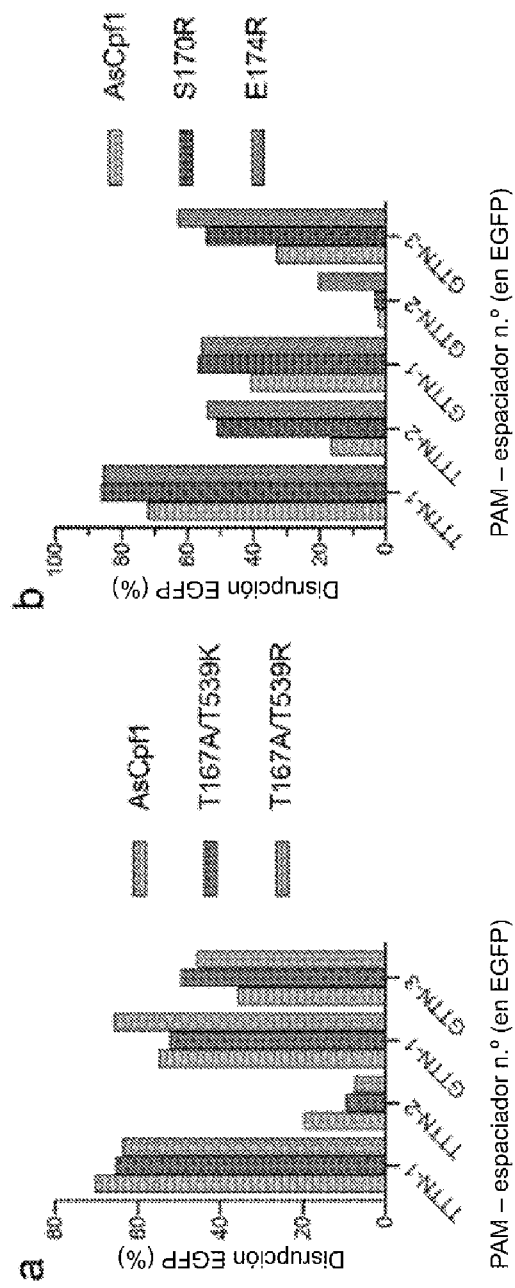


FIG. 5A

FIG. 5B

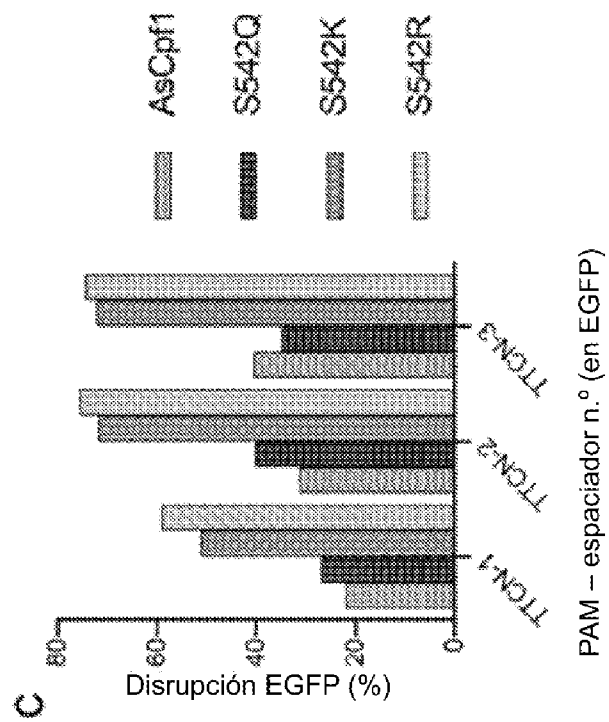


FIG. 5C

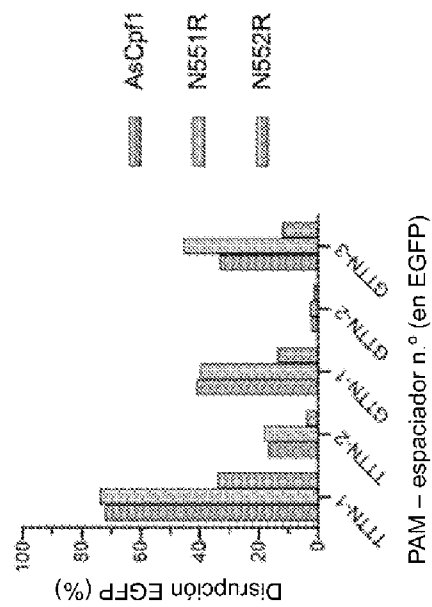


FIG. 5E

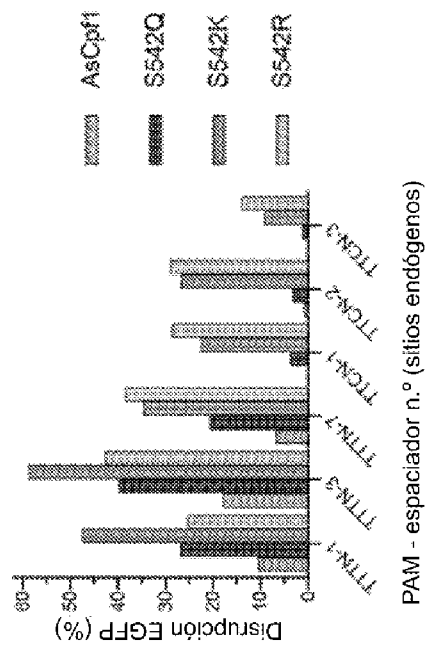


FIG. 5D

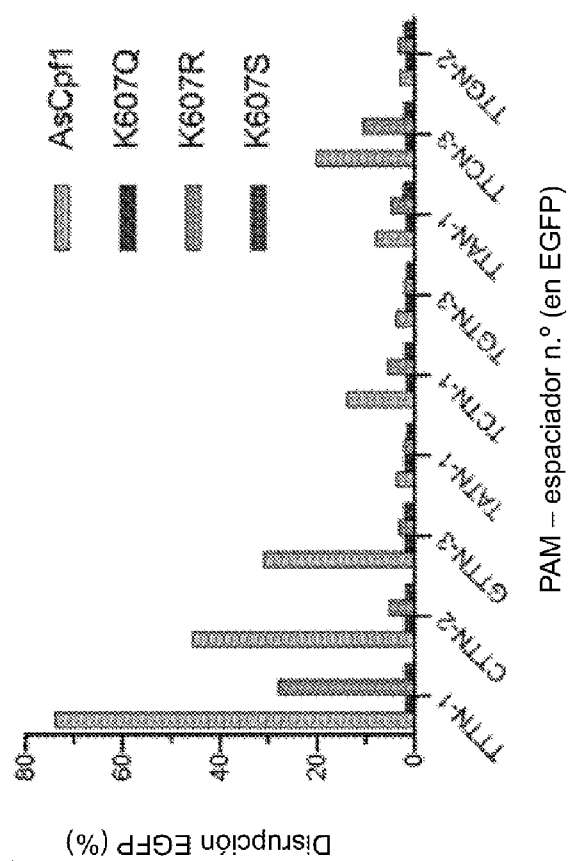
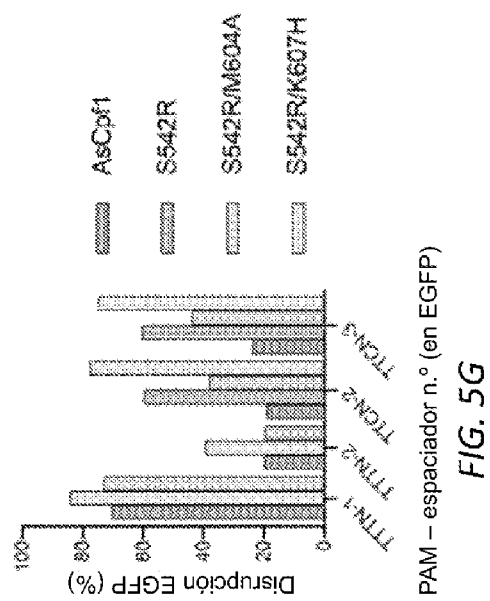
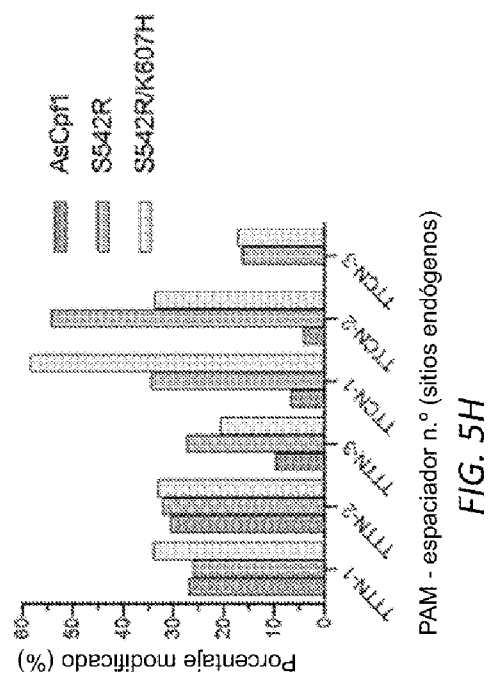


FIG. 5F



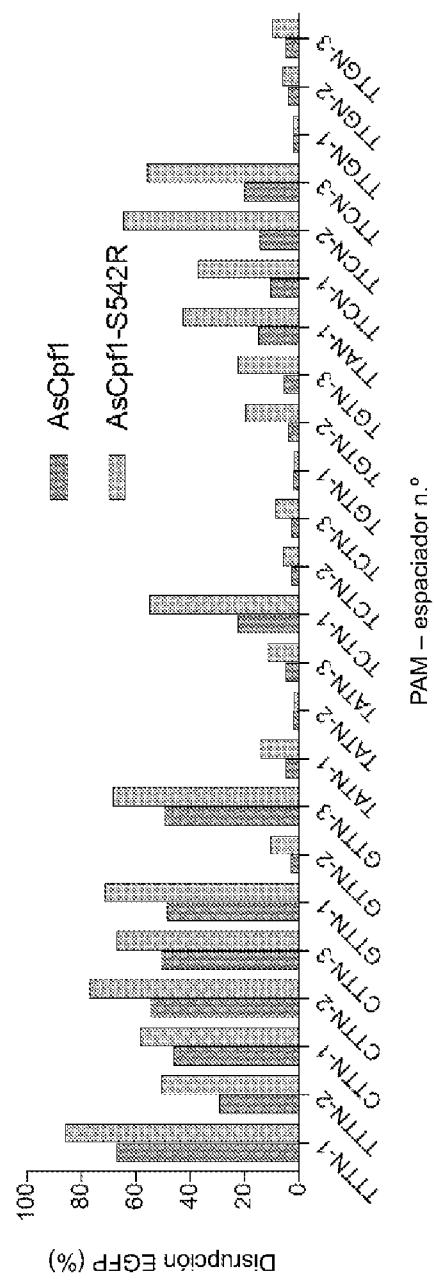


FIG. 6A

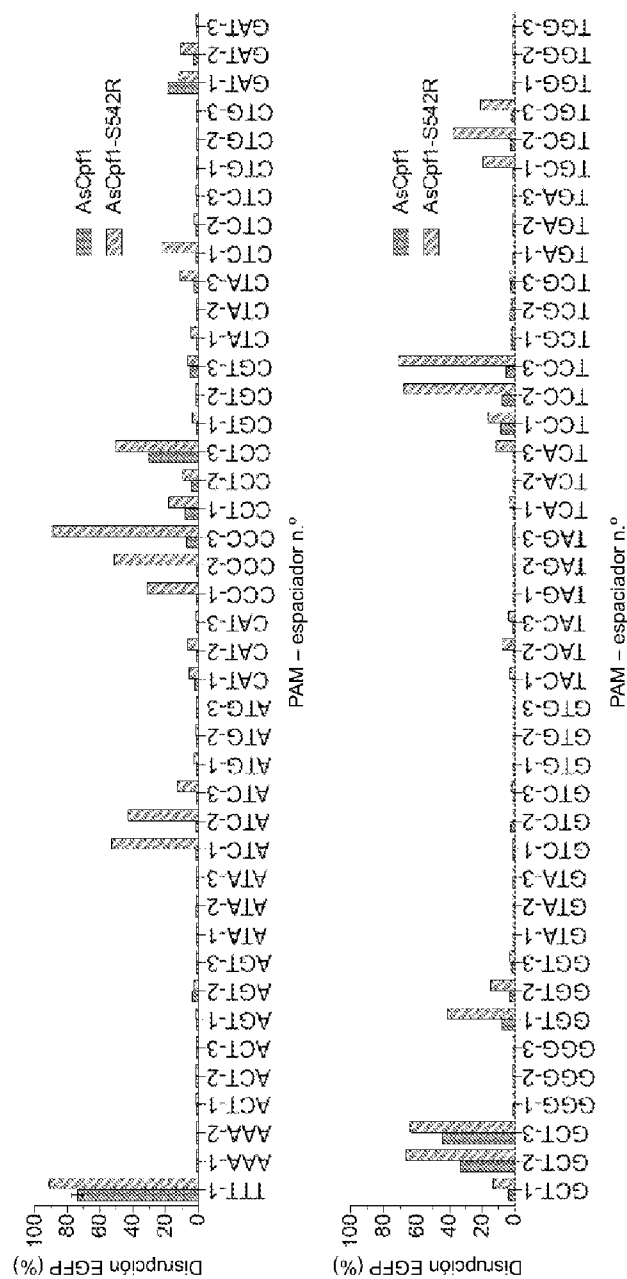
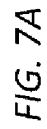


FIG. 6B



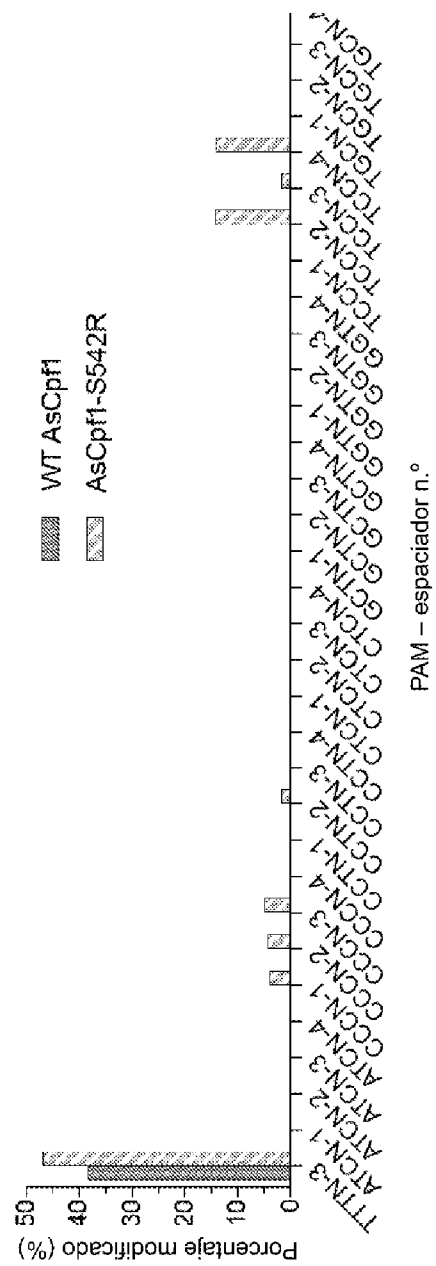
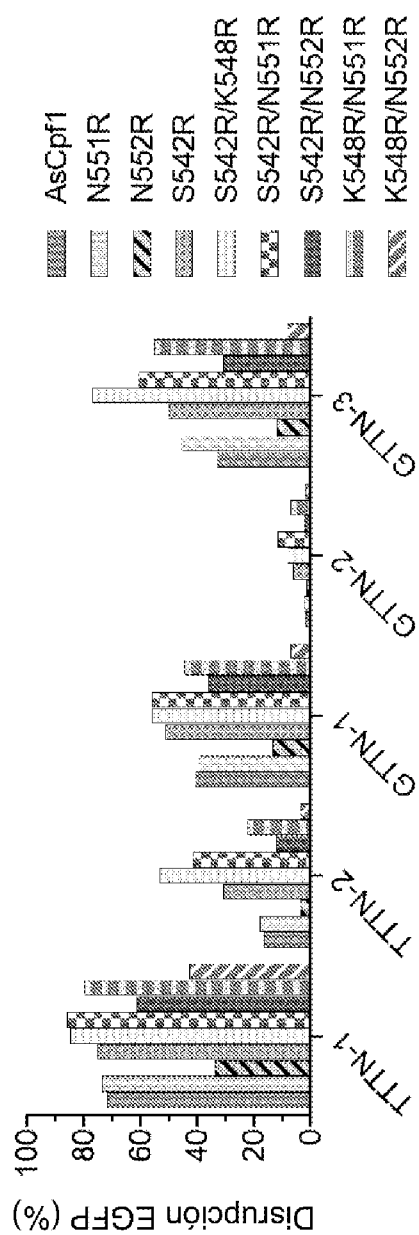
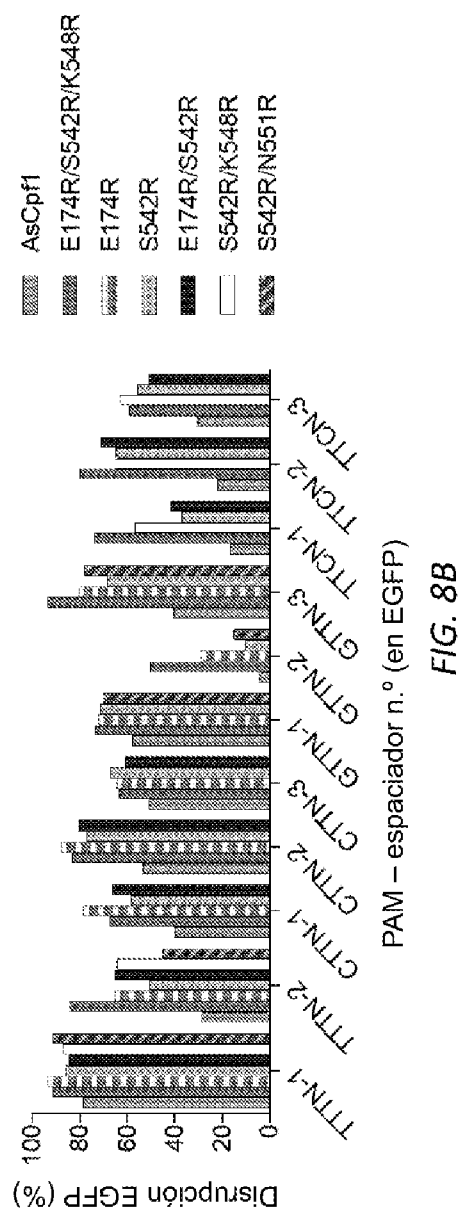


FIG. 7B



PAM – espaciador n.º (en EGFP)

FIG. 8A



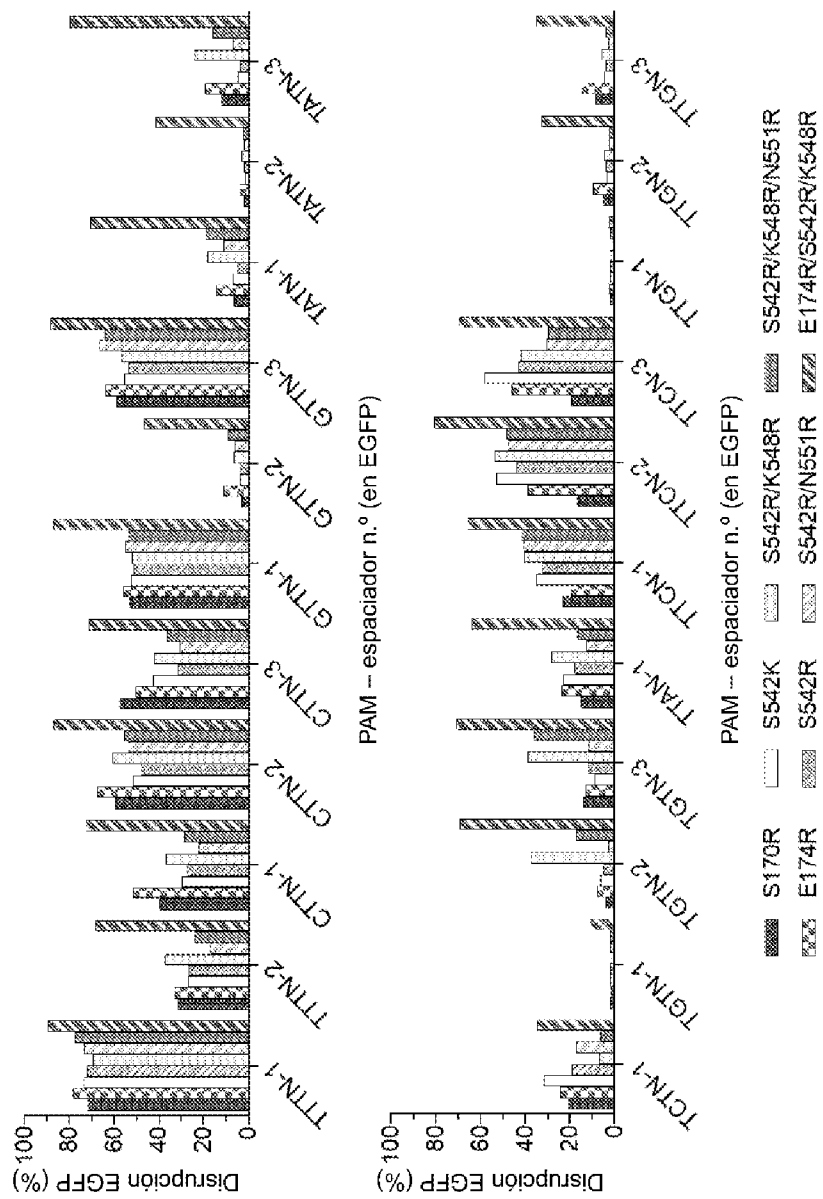
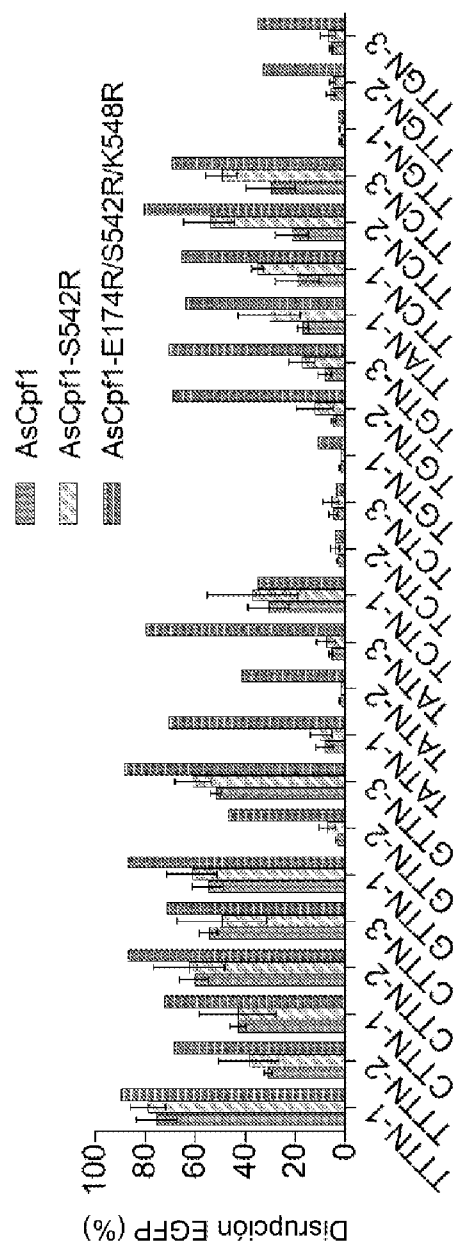


FIG. 8C



PAM – espaciador n.º

FIG. 9A

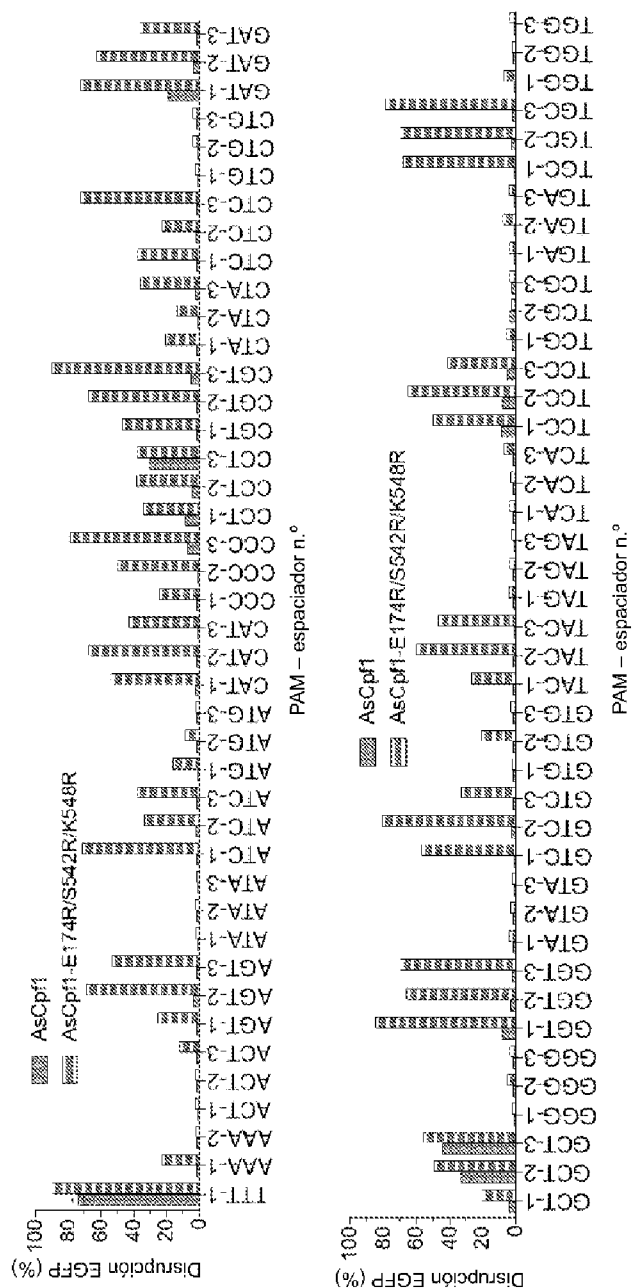


FIG. 9B



PAM - espaciador n.º (sitios endógenos)

FIG. 10A

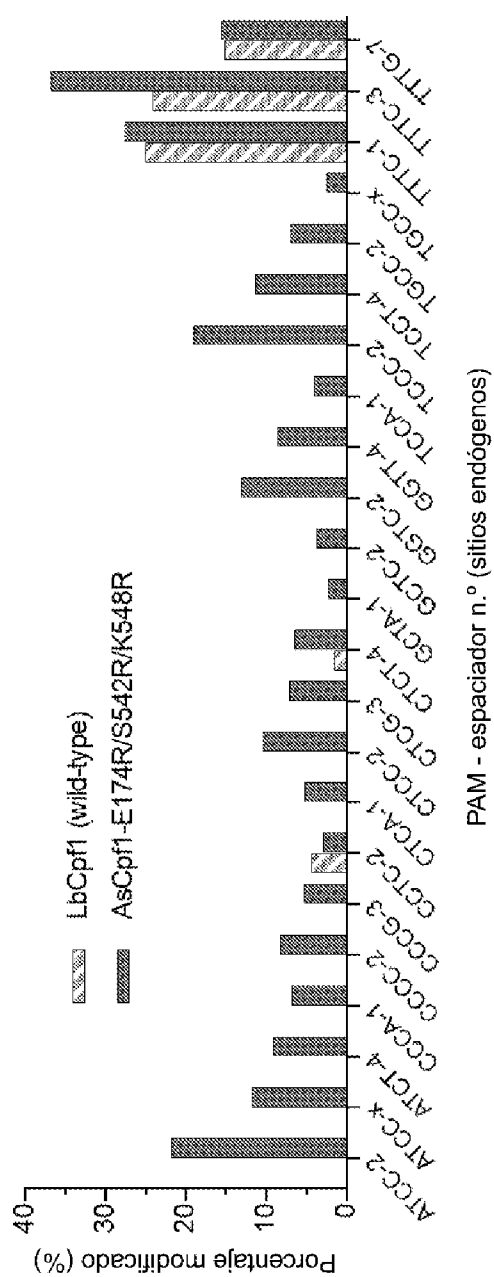


FIG. 10B

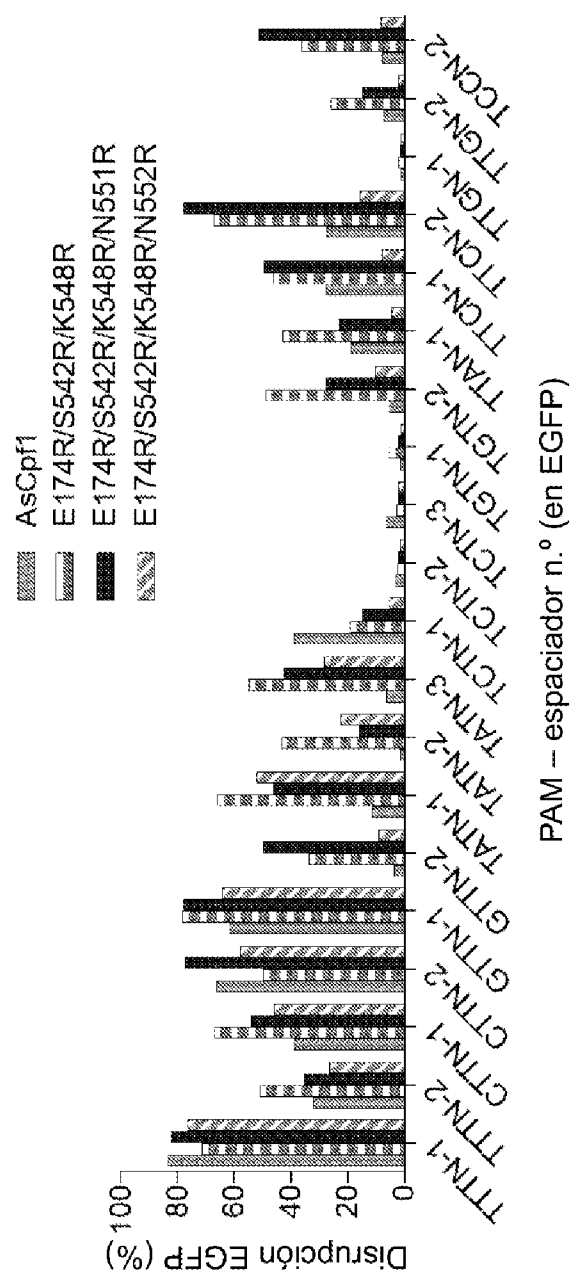
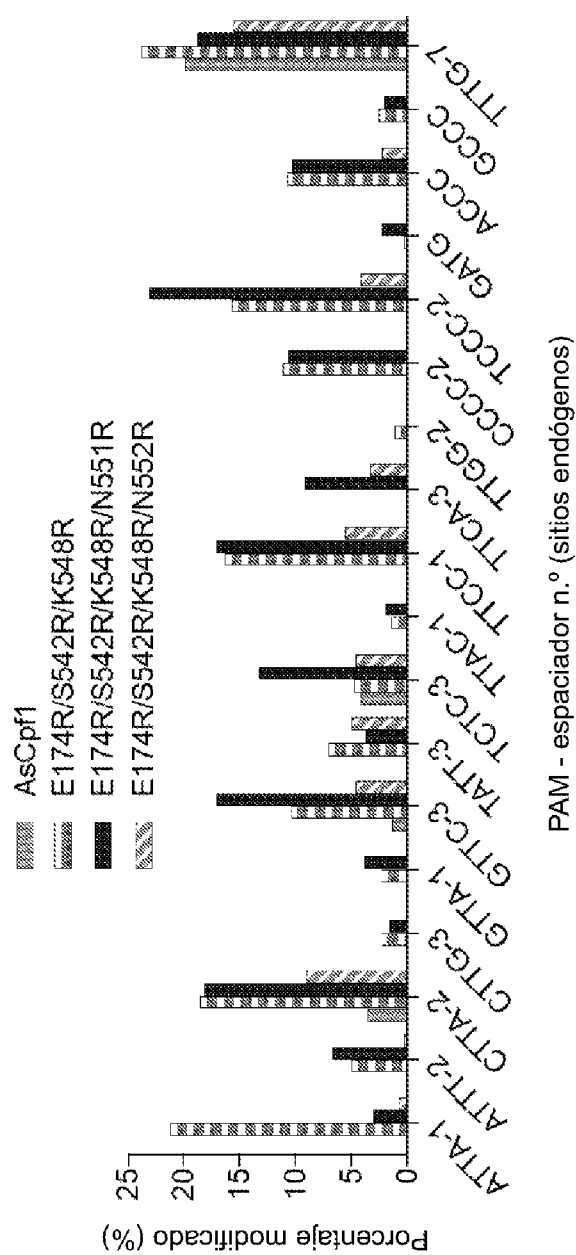
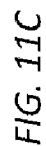


FIG. 11A





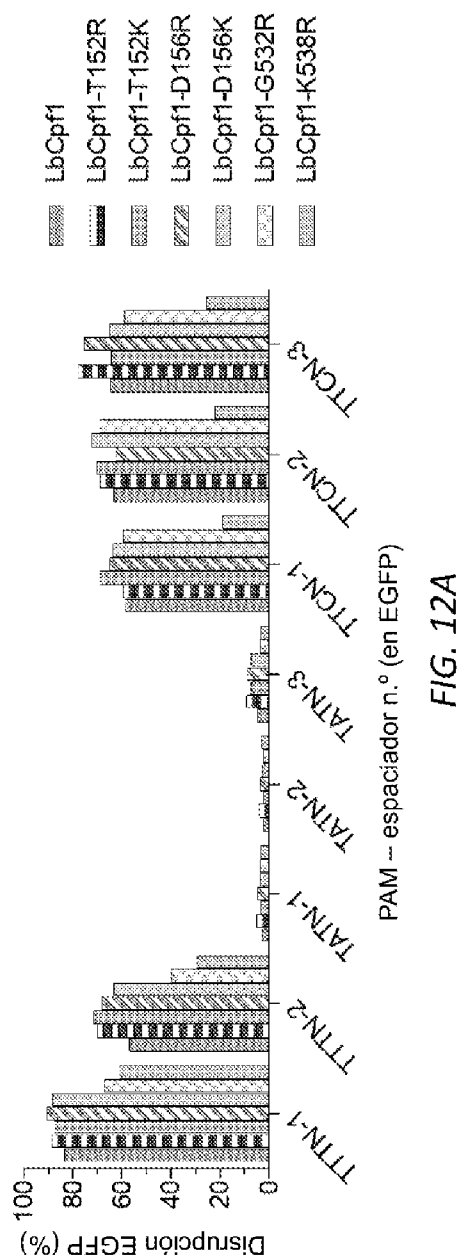


FIG. 12A

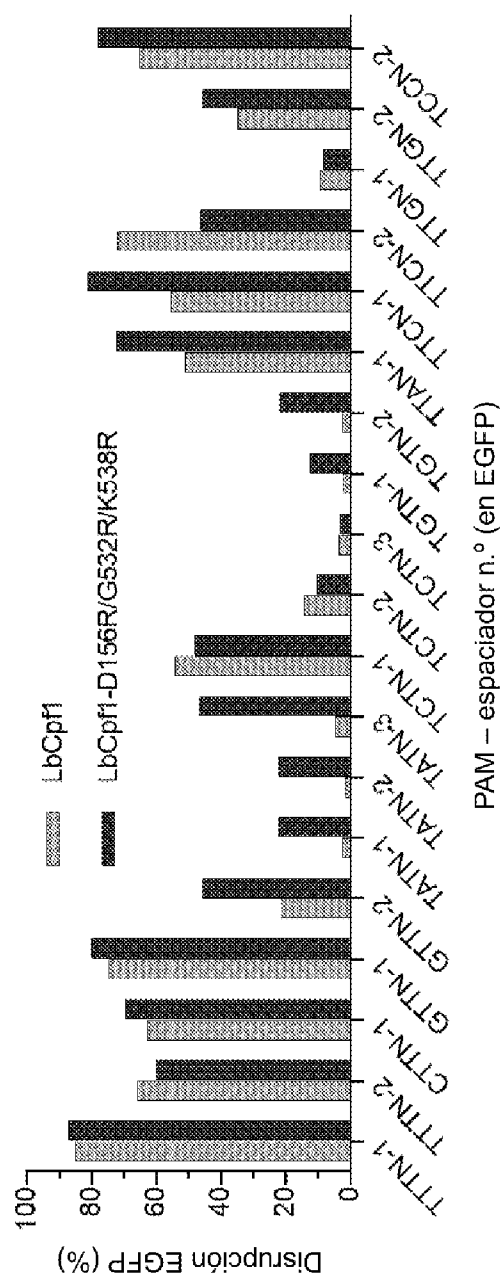


FIG. 12B

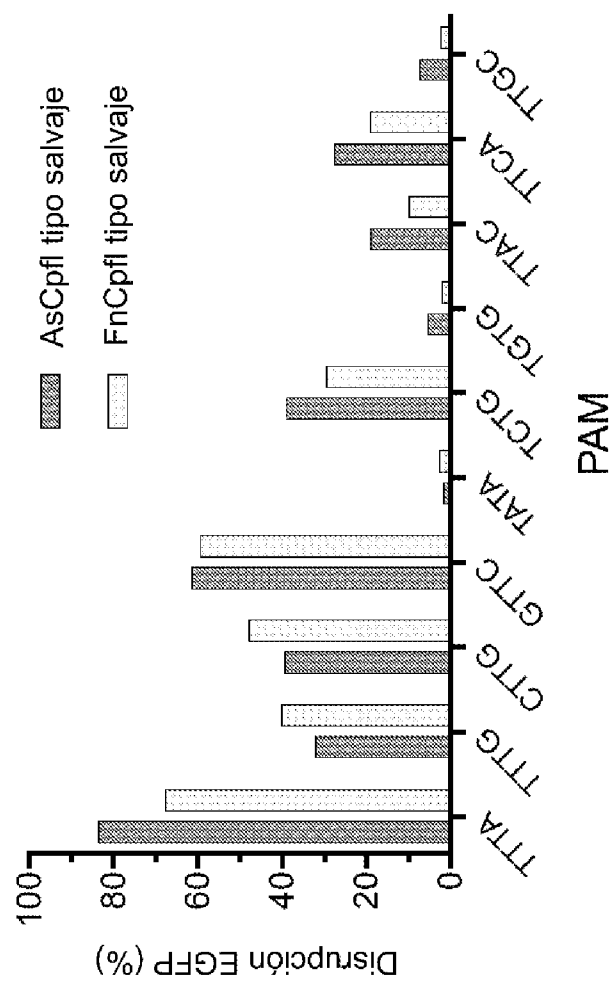
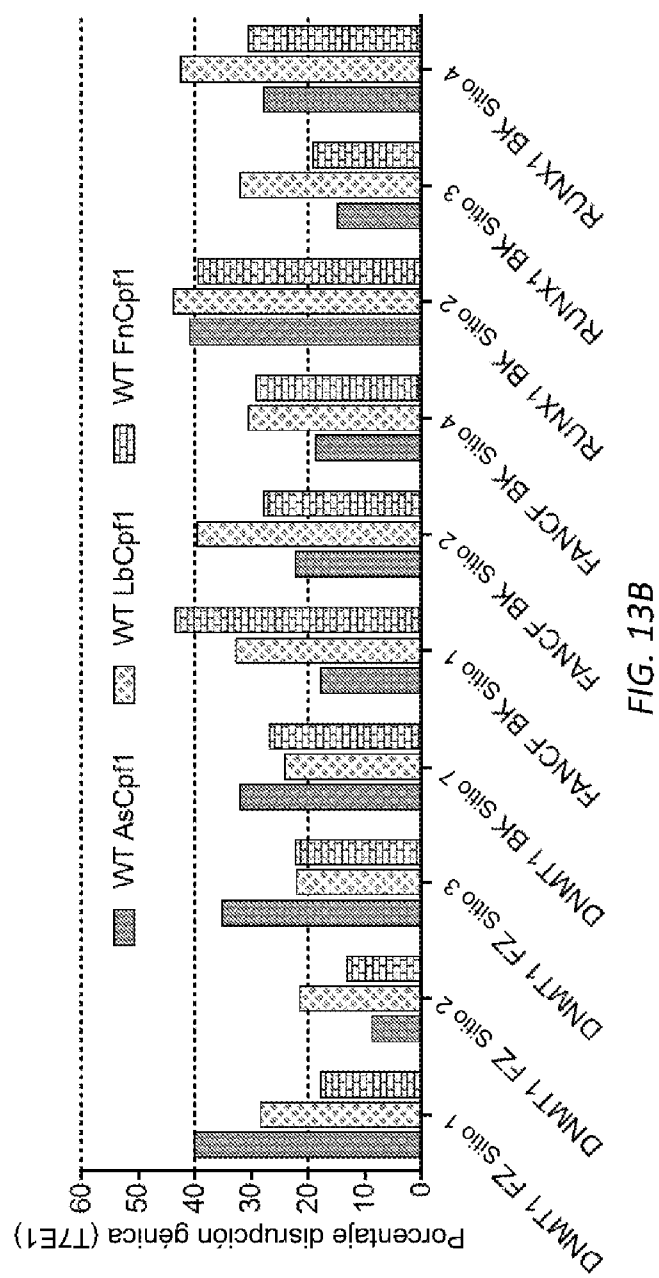


FIG. 13A



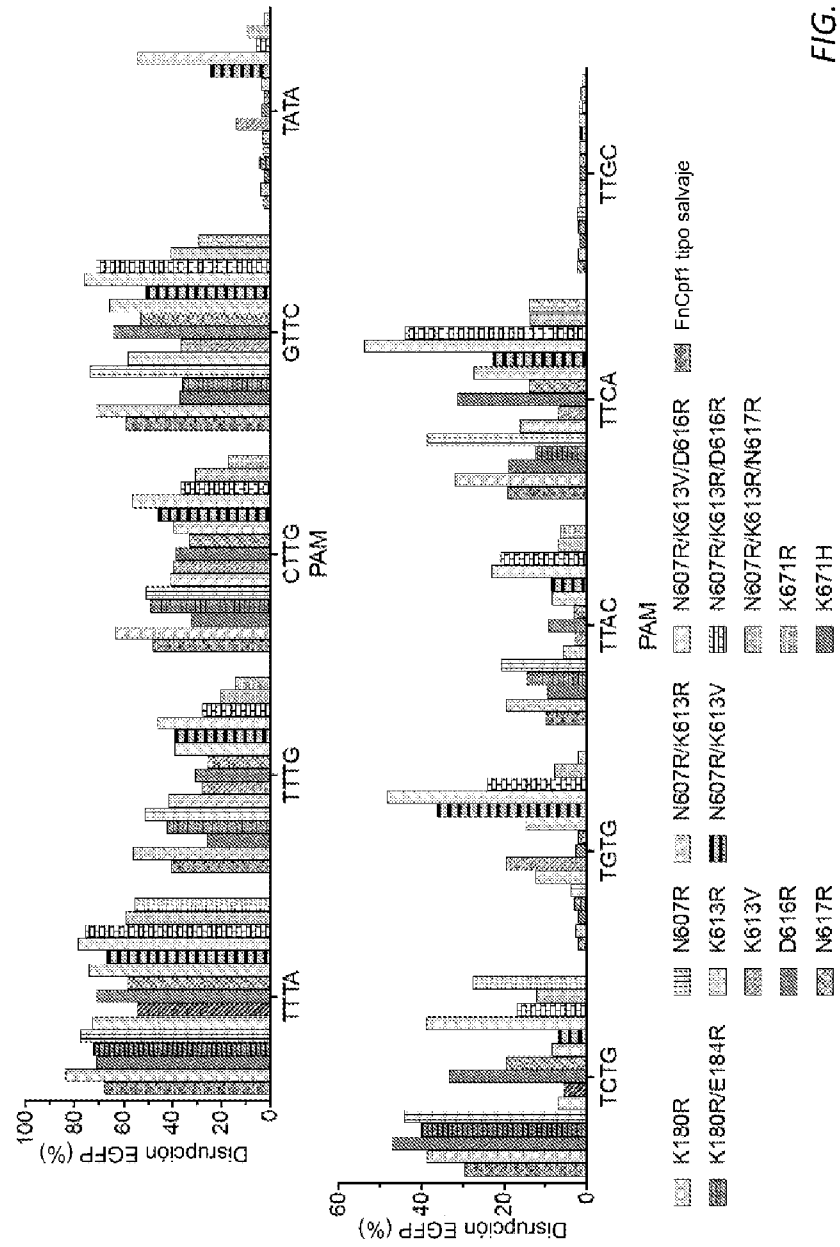
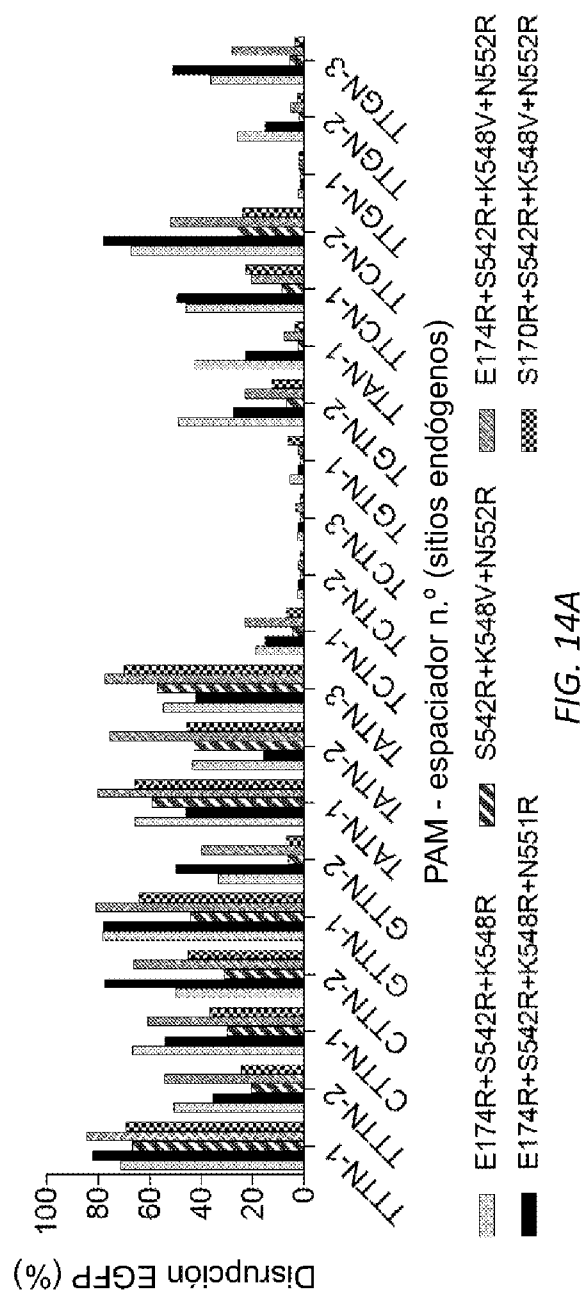


FIG. 13C



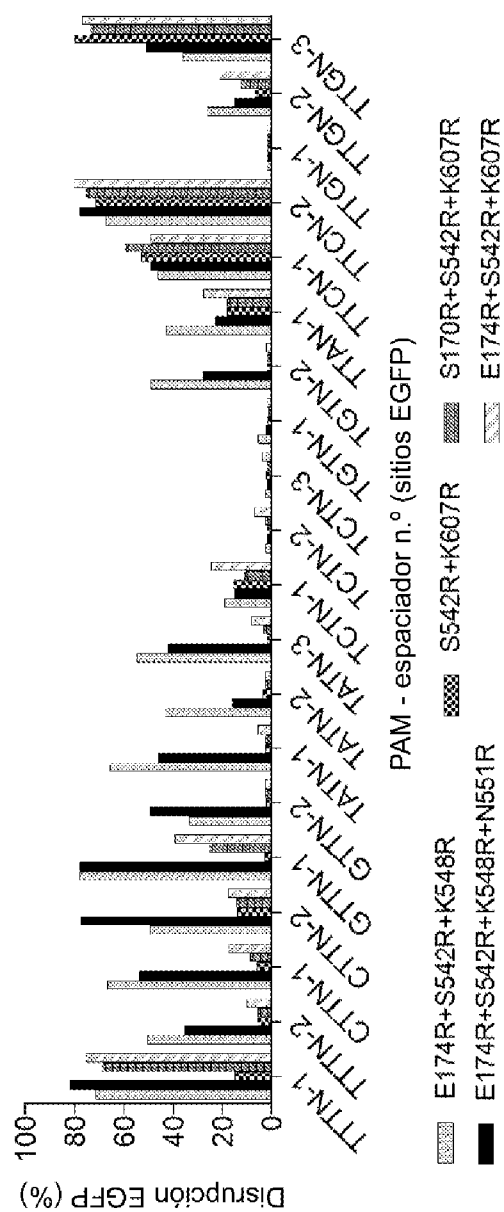


FIG. 14B

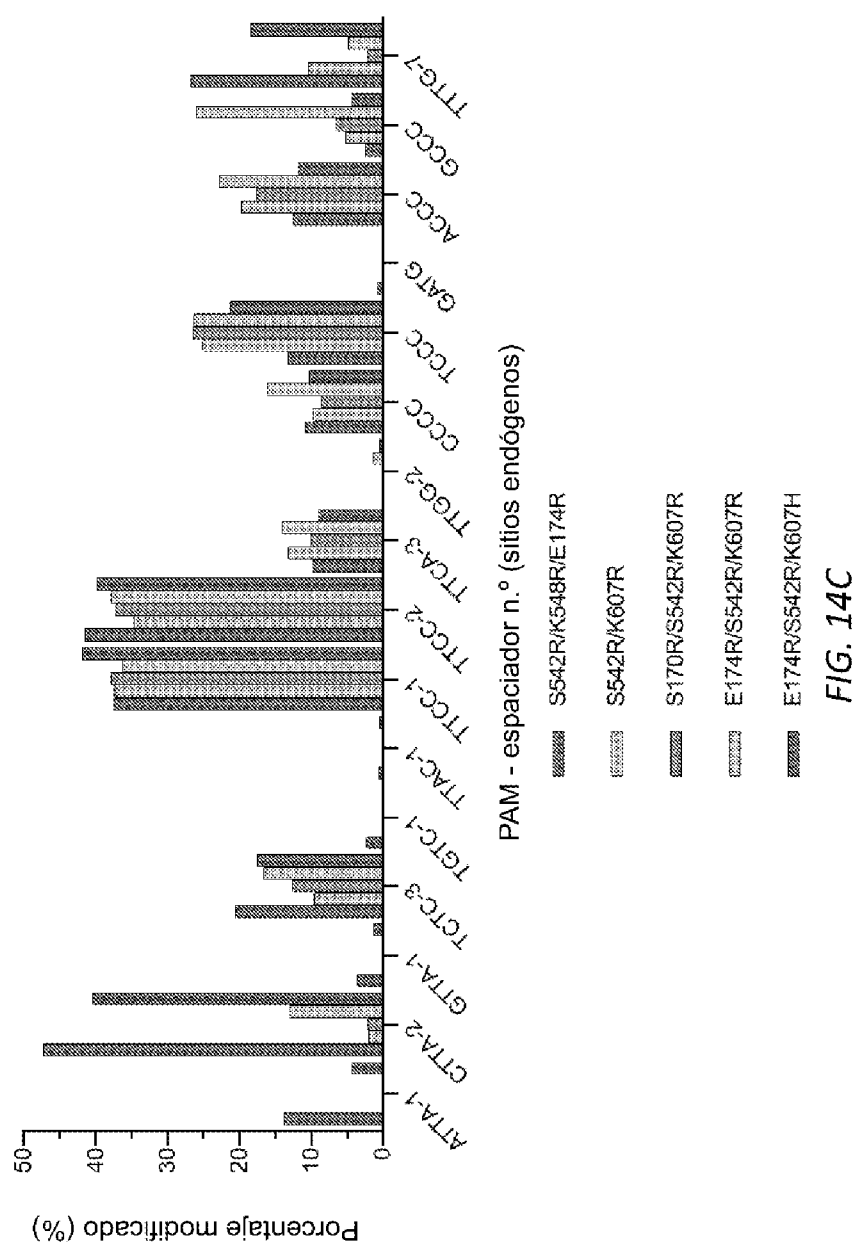




FIG. 15A

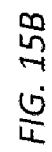




FIG. 15C

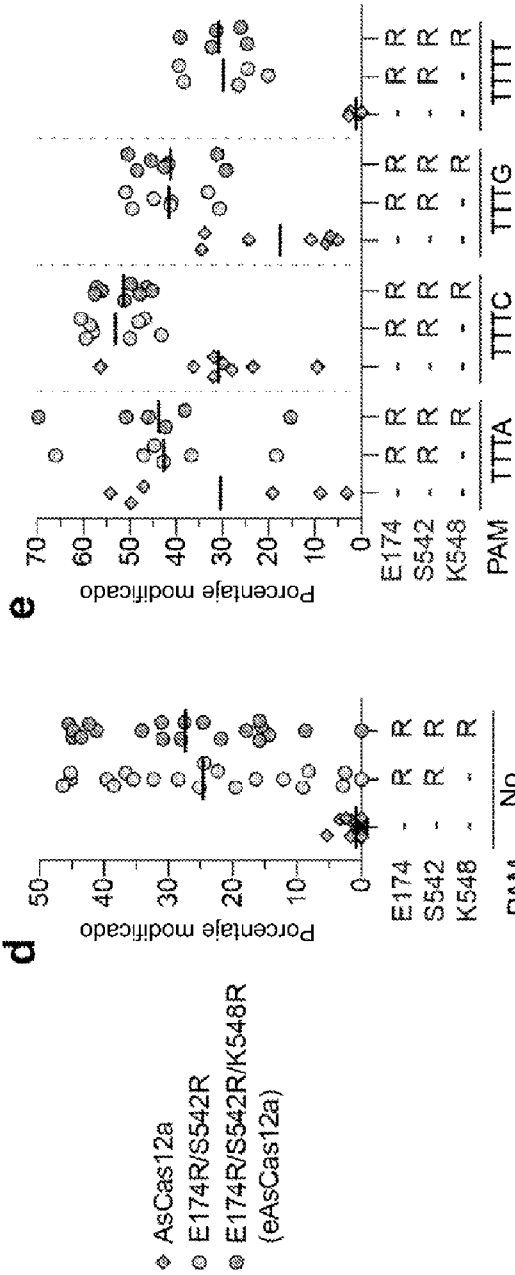
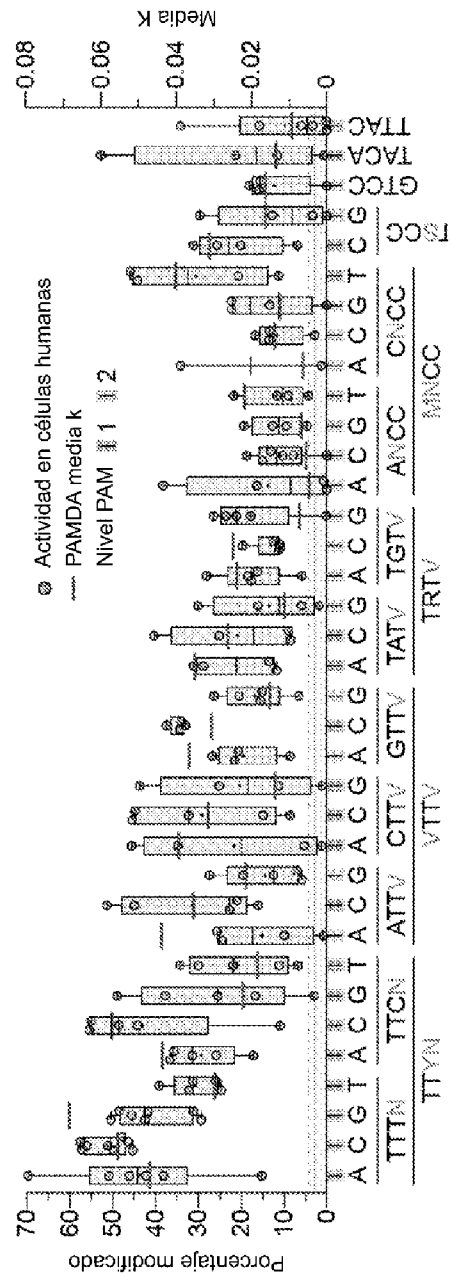


FIG. 15E

FIG. 15D



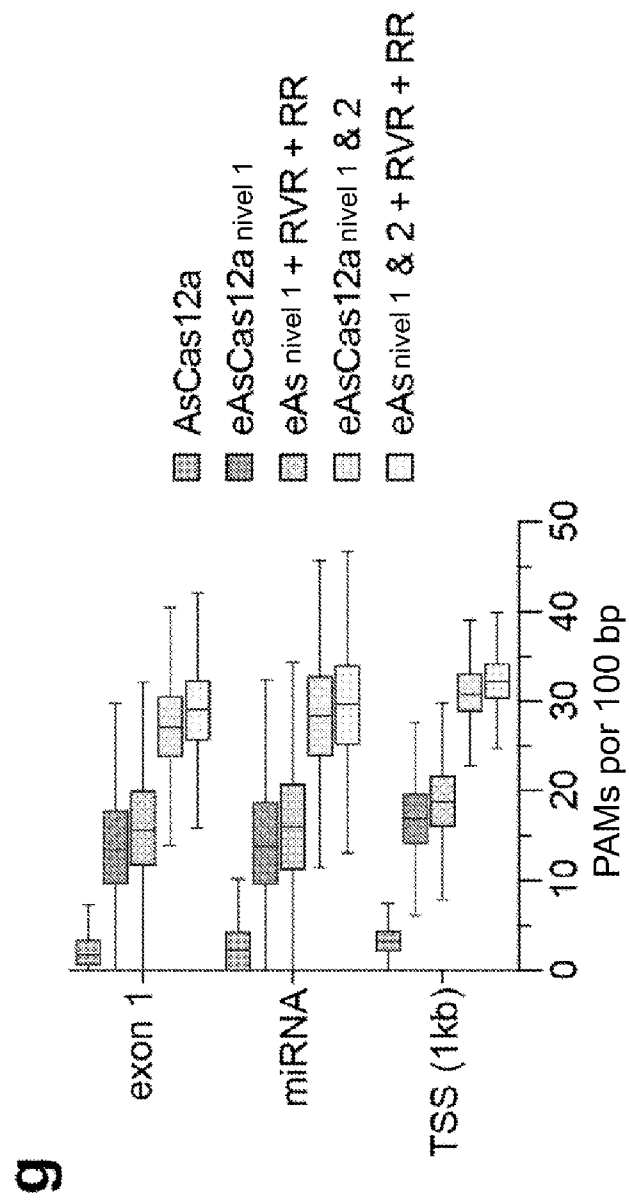


FIG. 15G

EP 3 612 551 B1

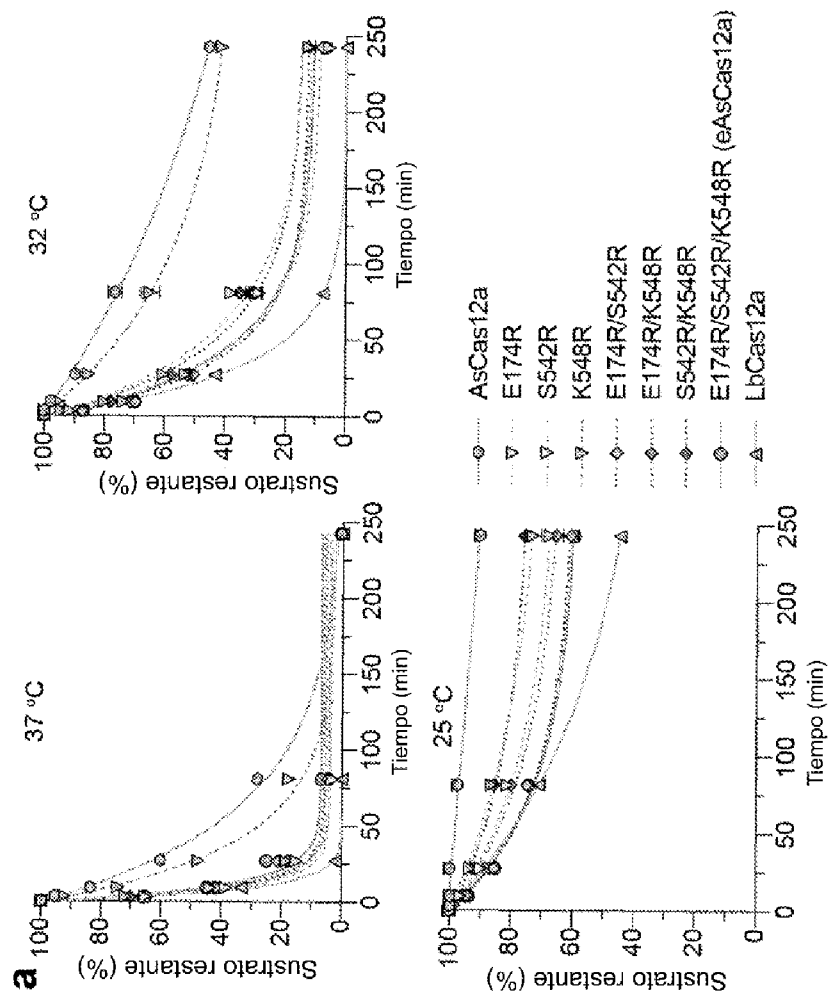


FIG. 16A

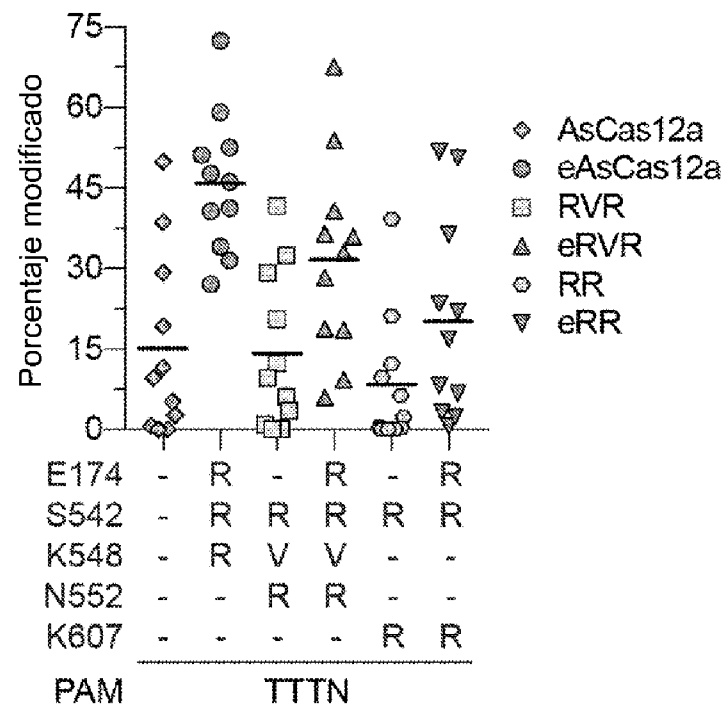
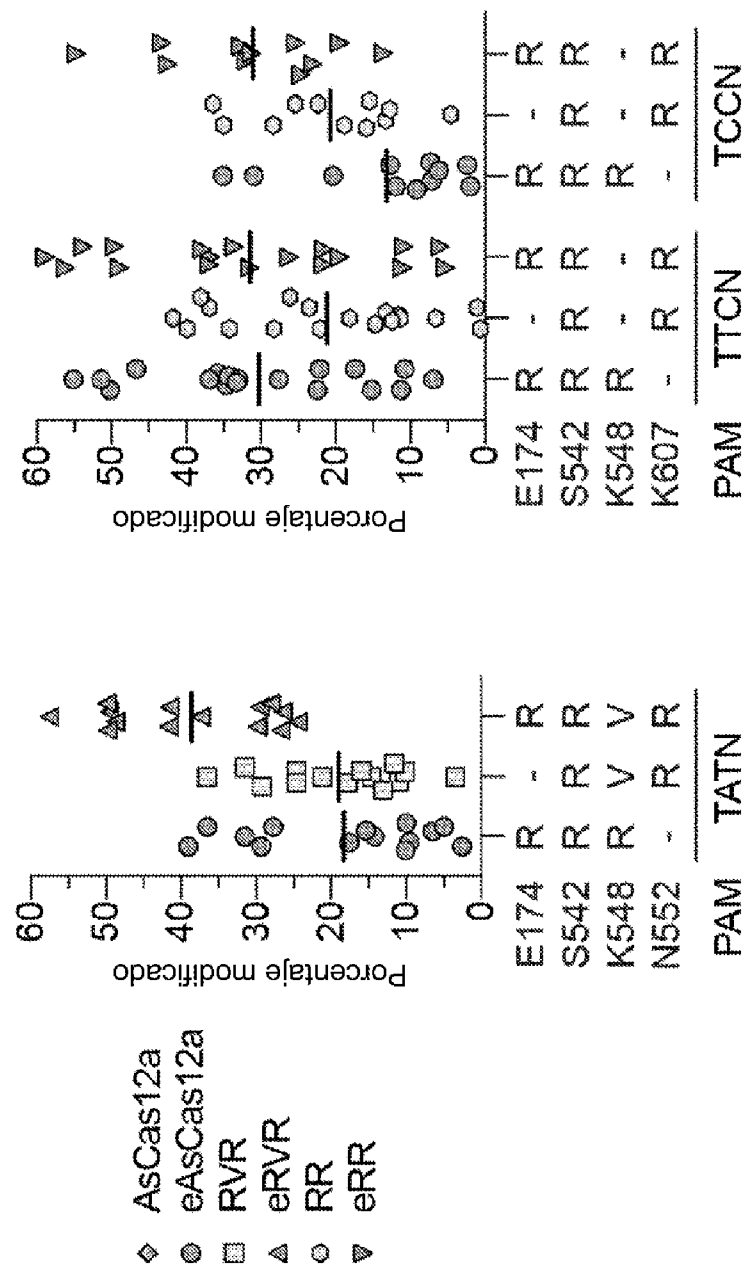


FIG. 16B



FIGs. 16C-D

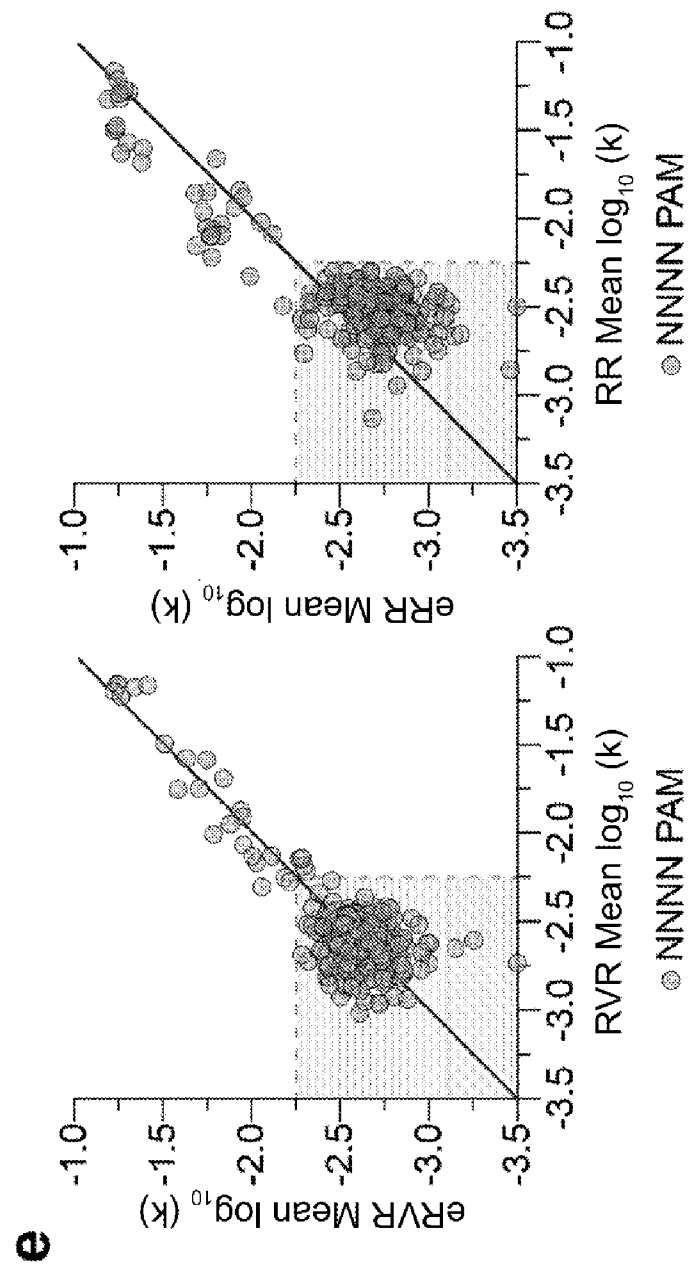
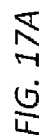
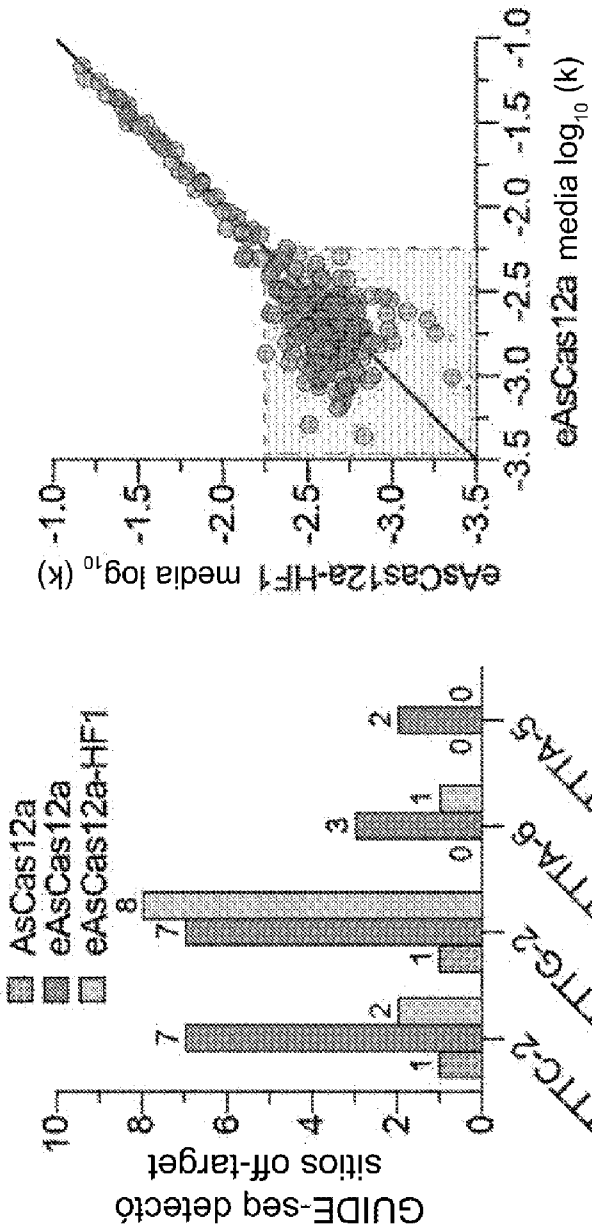


FIG. 16E





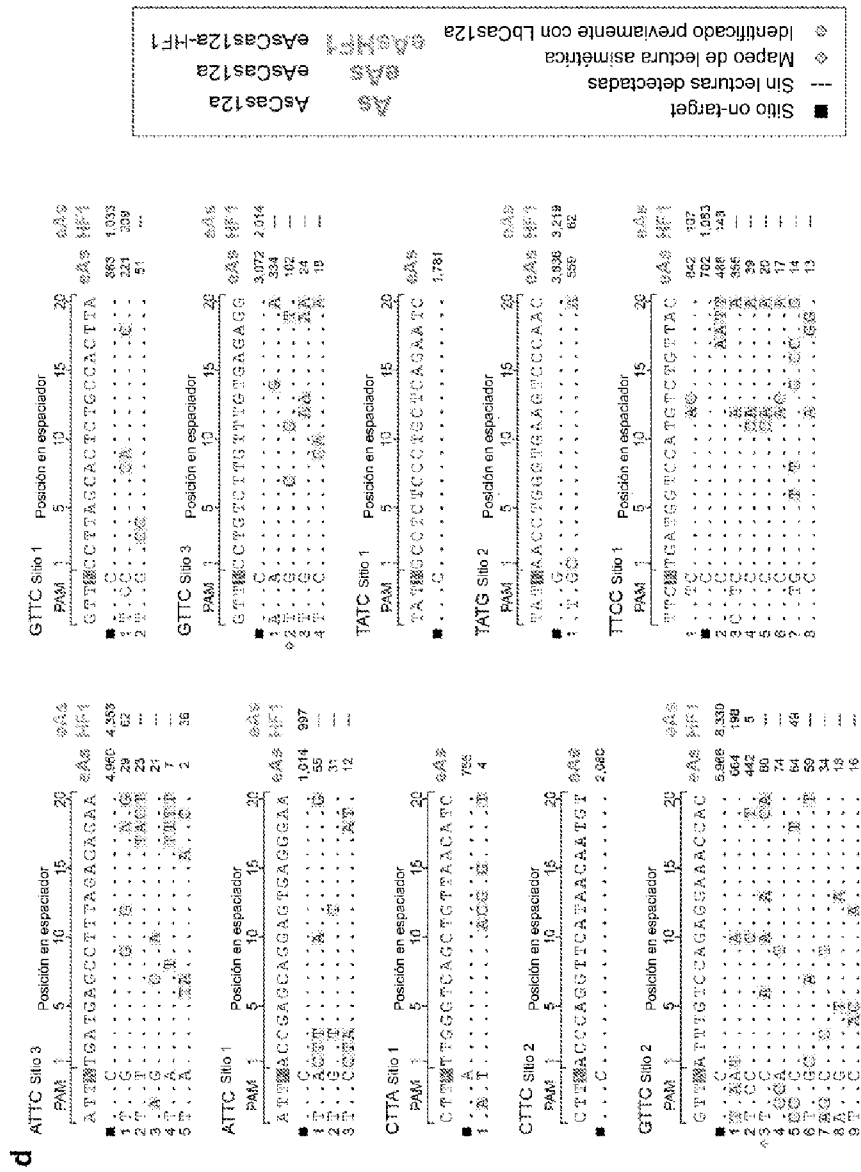
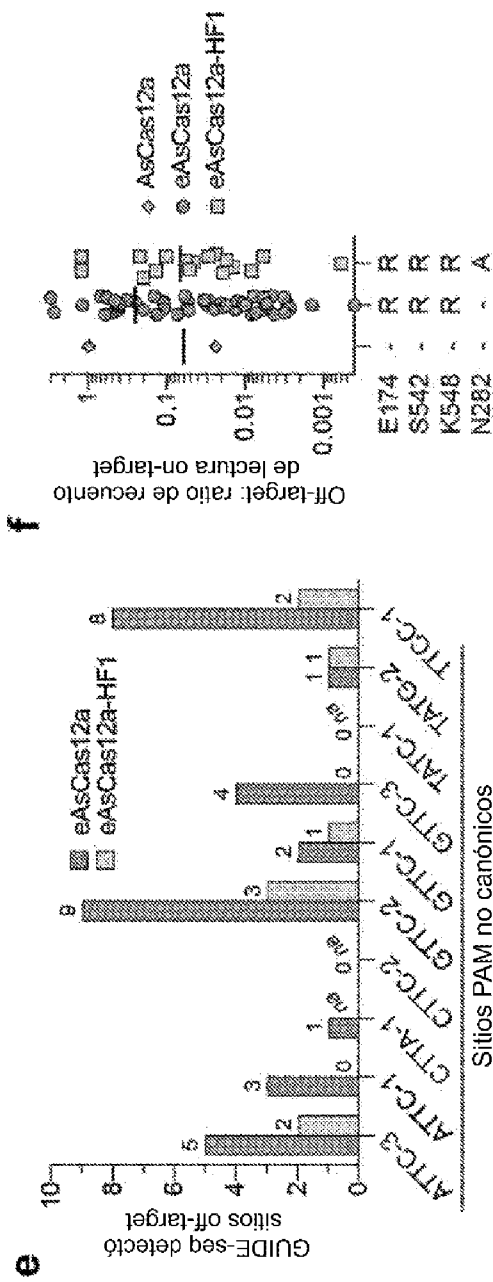
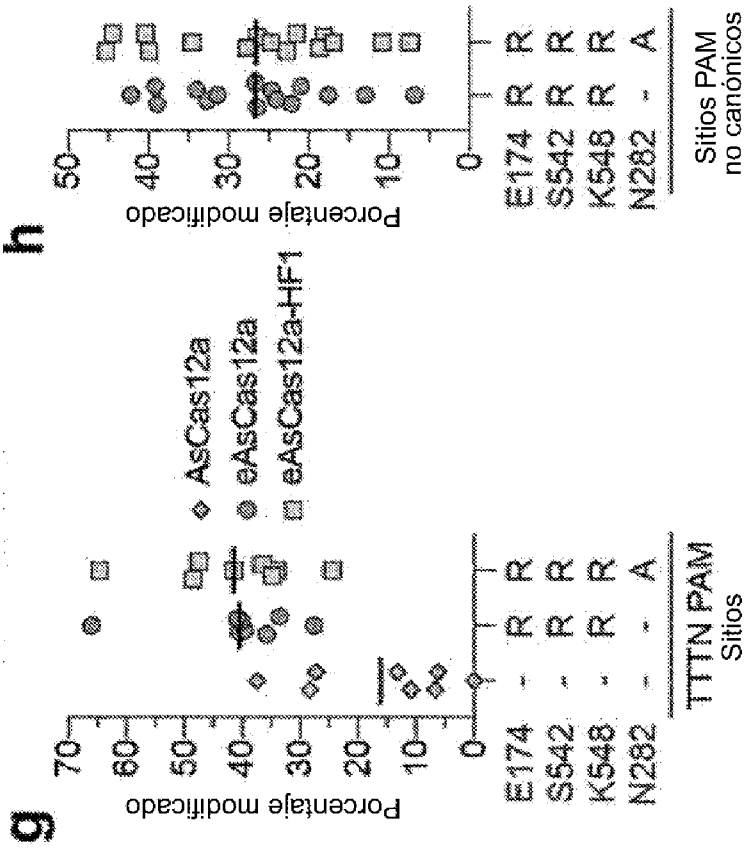


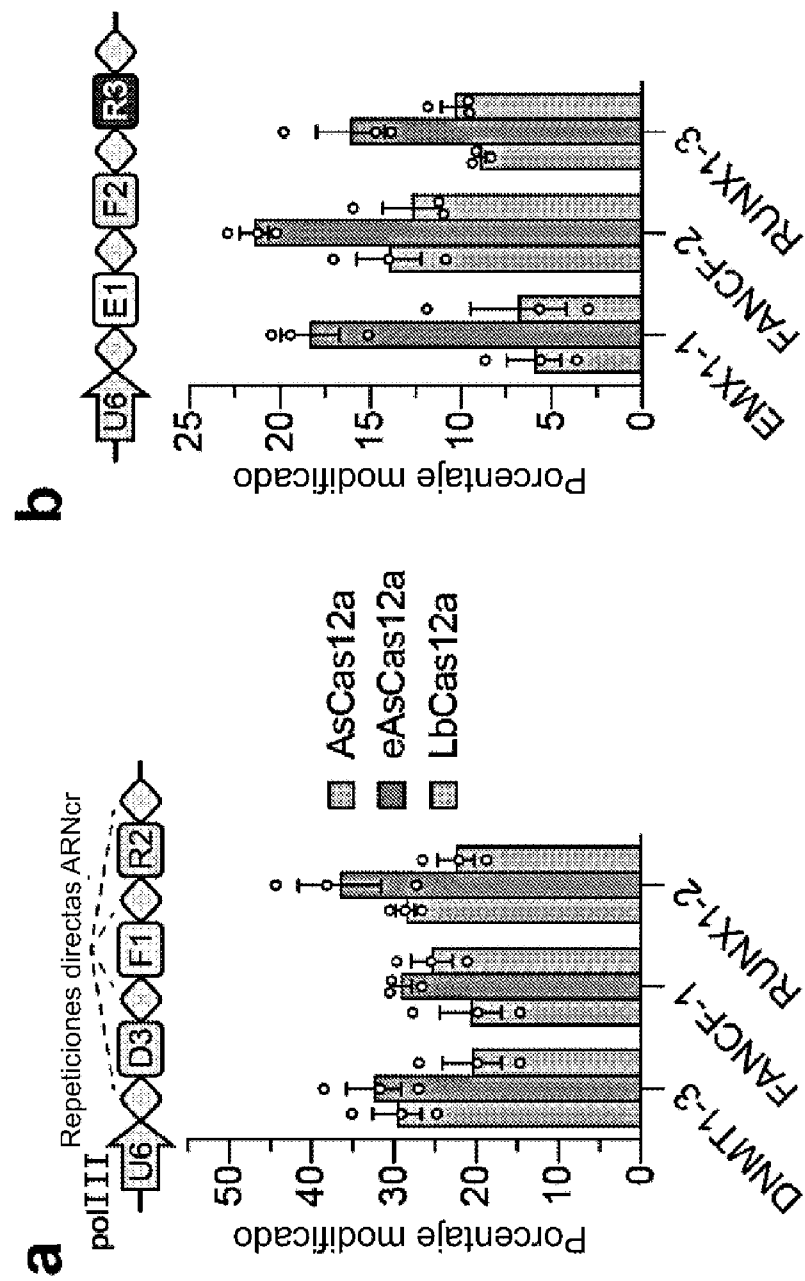
FIG. 17D



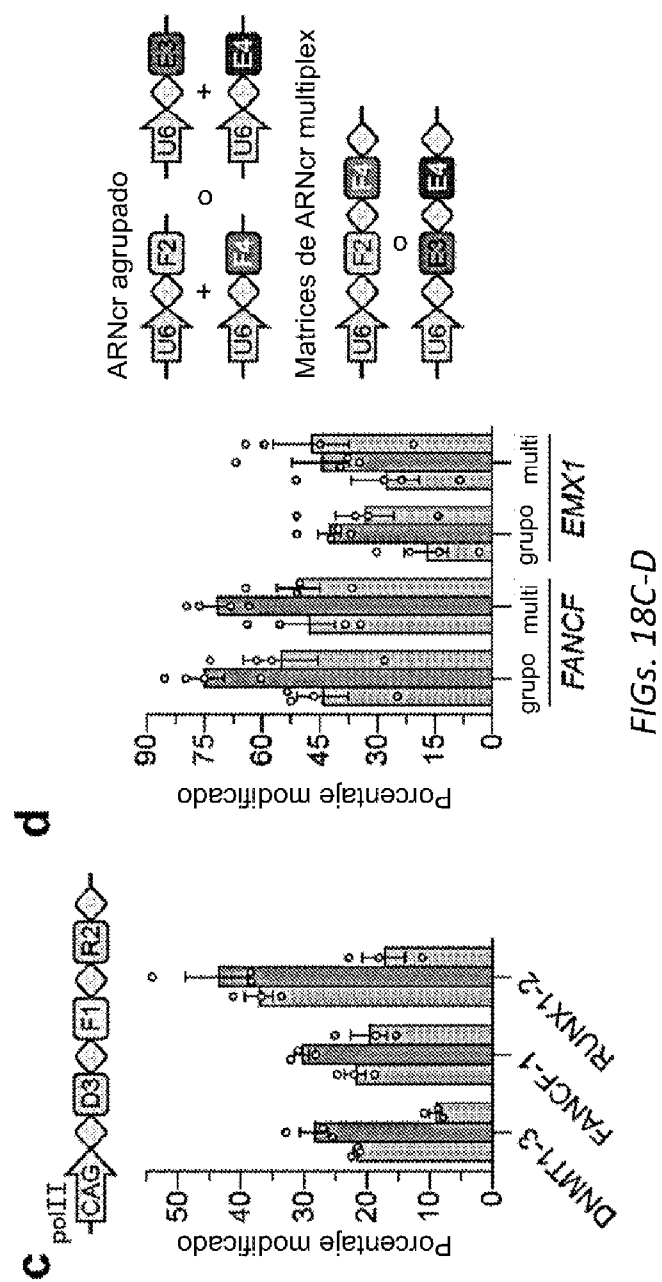
FIGs. 17E-F



FIGs. 17G-H



FIGs. 18A-B



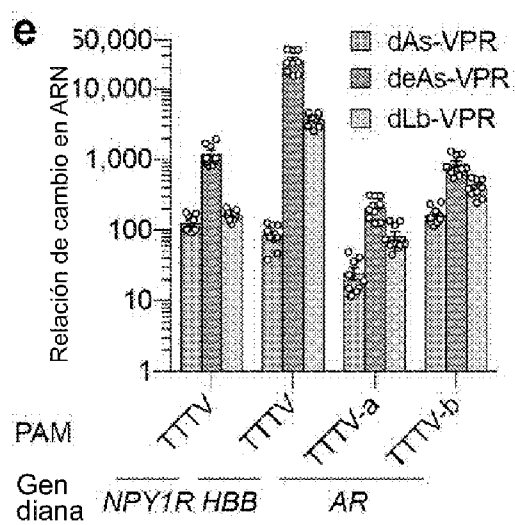
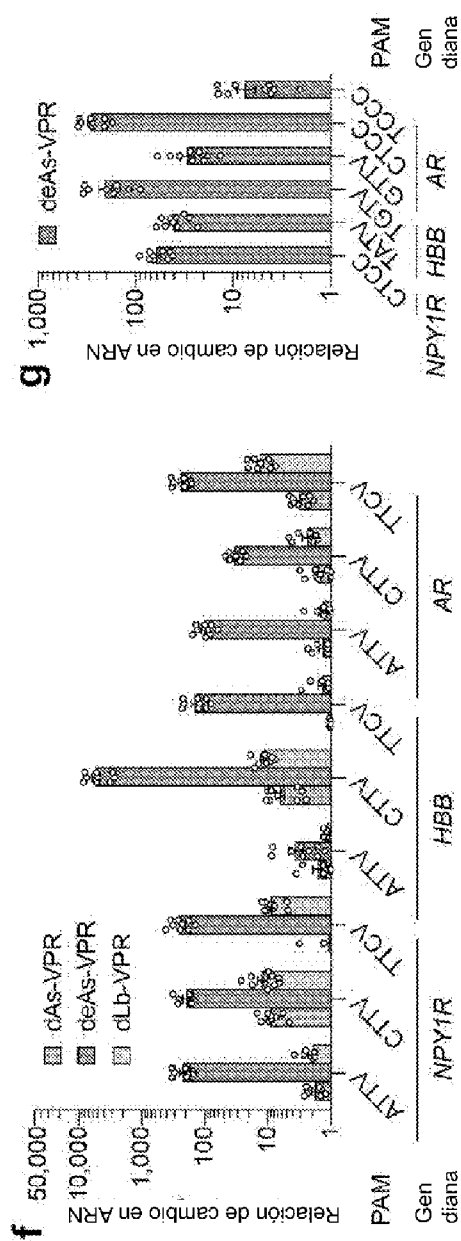


FIG. 18E



FIGs. 18F-G

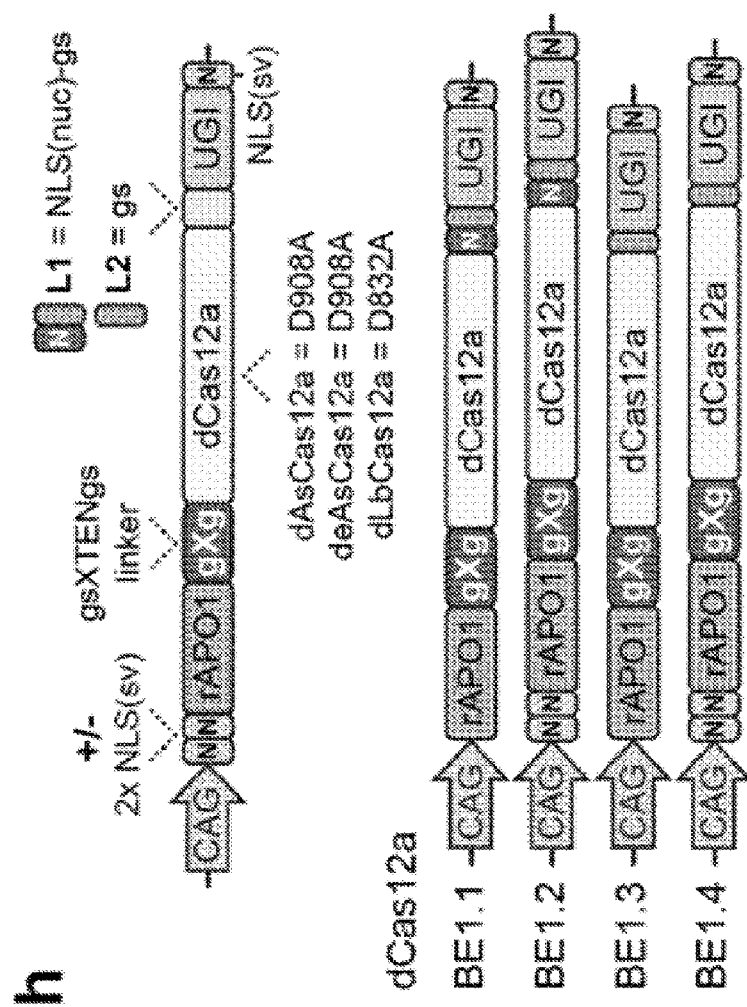


FIG. 18H

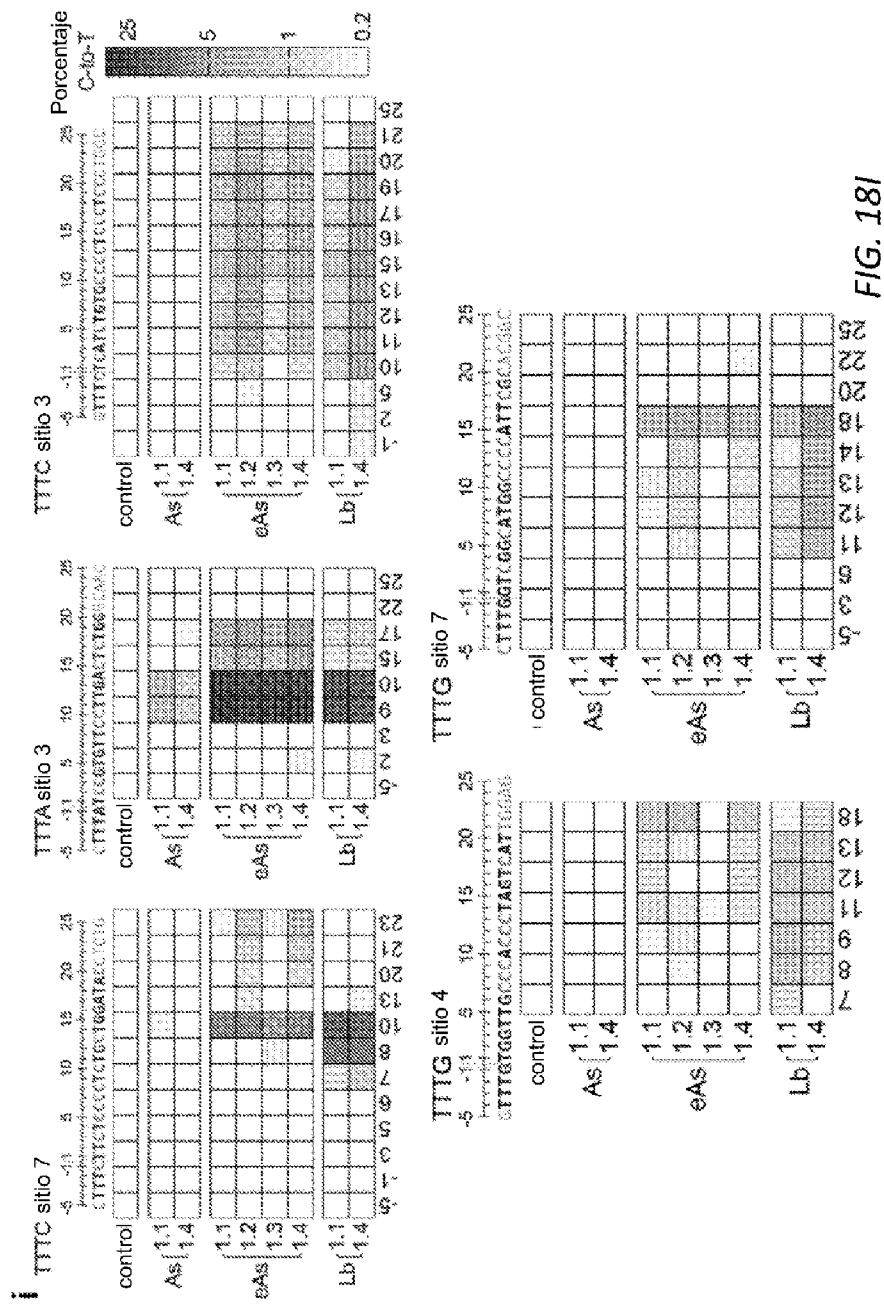


FIG. 18I

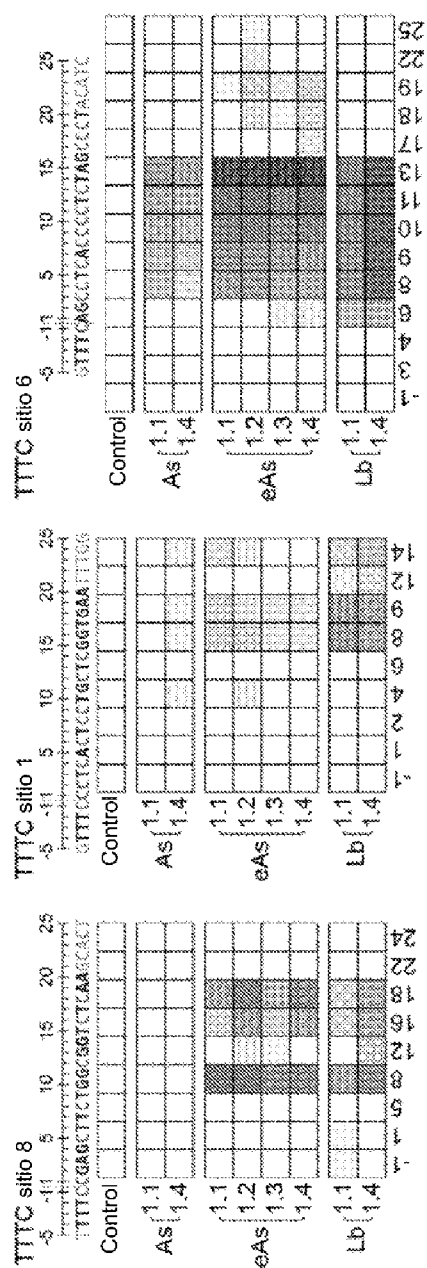


FIG. 18I (continuado)

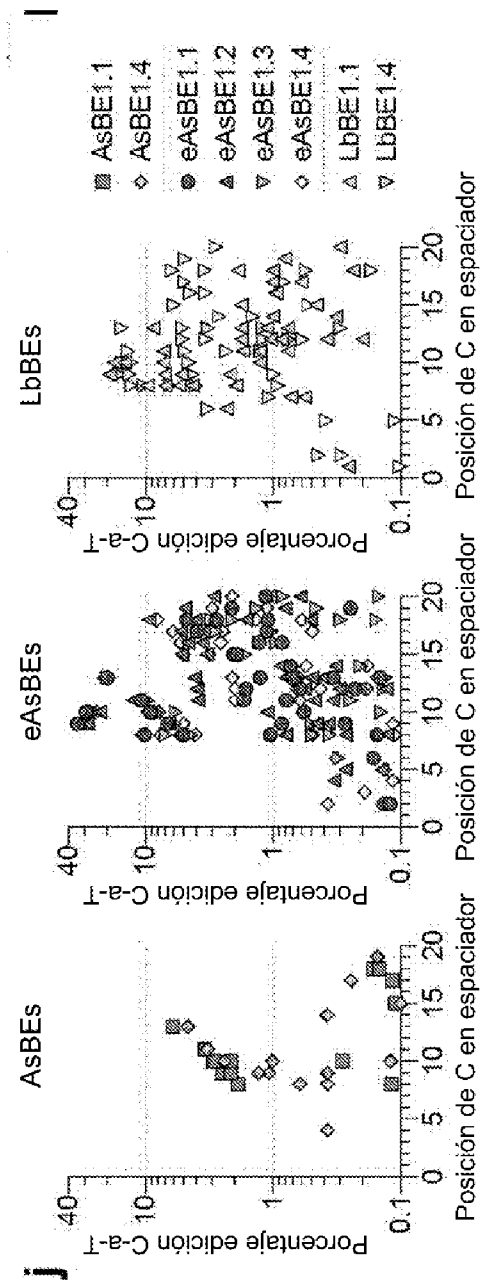
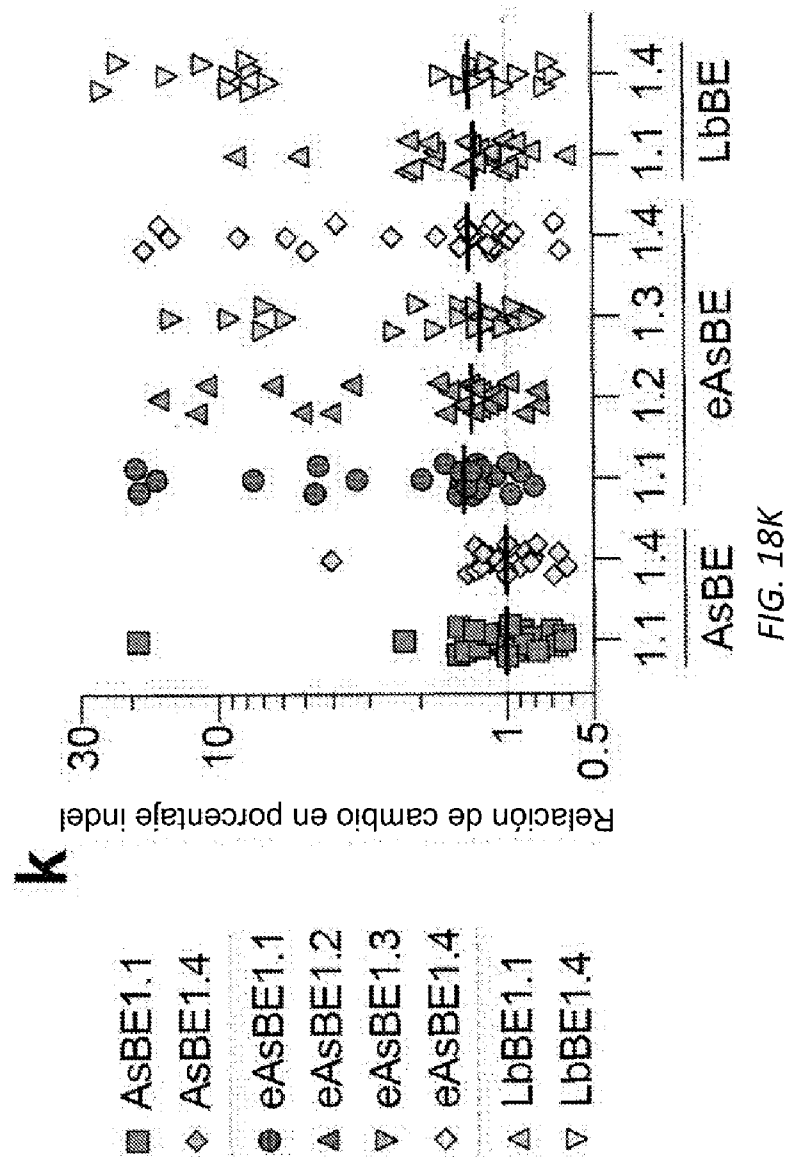
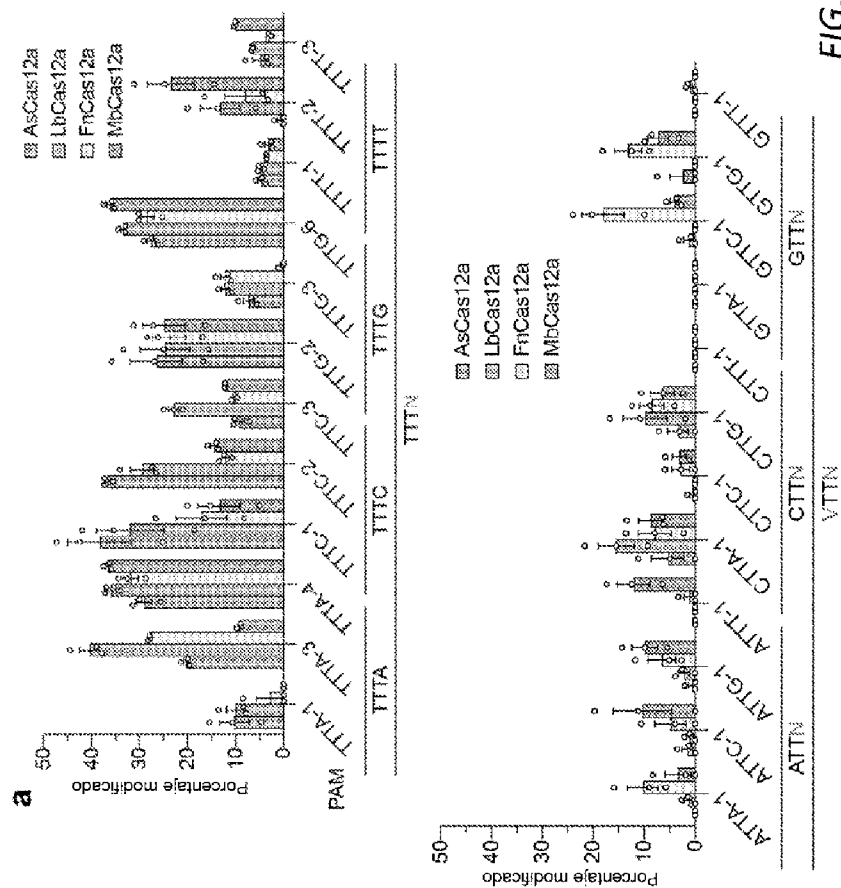


FIG. 18J





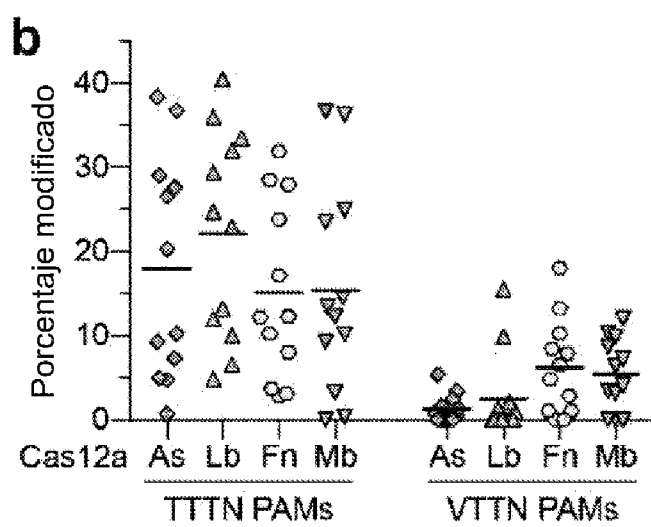


FIG. 19B

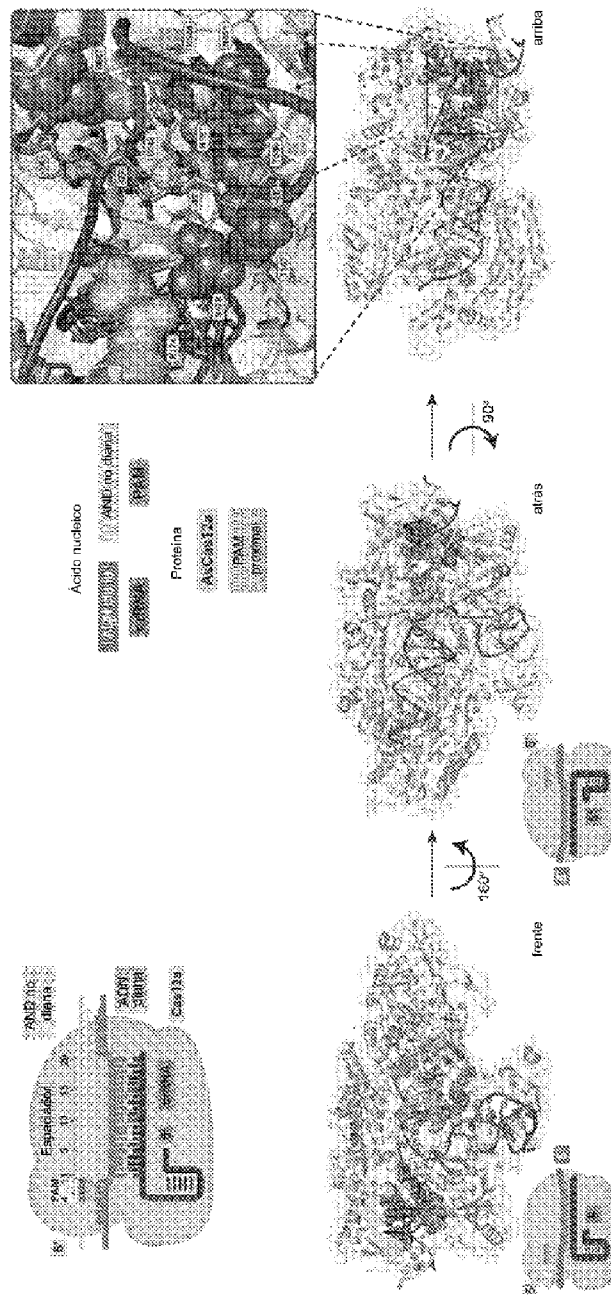
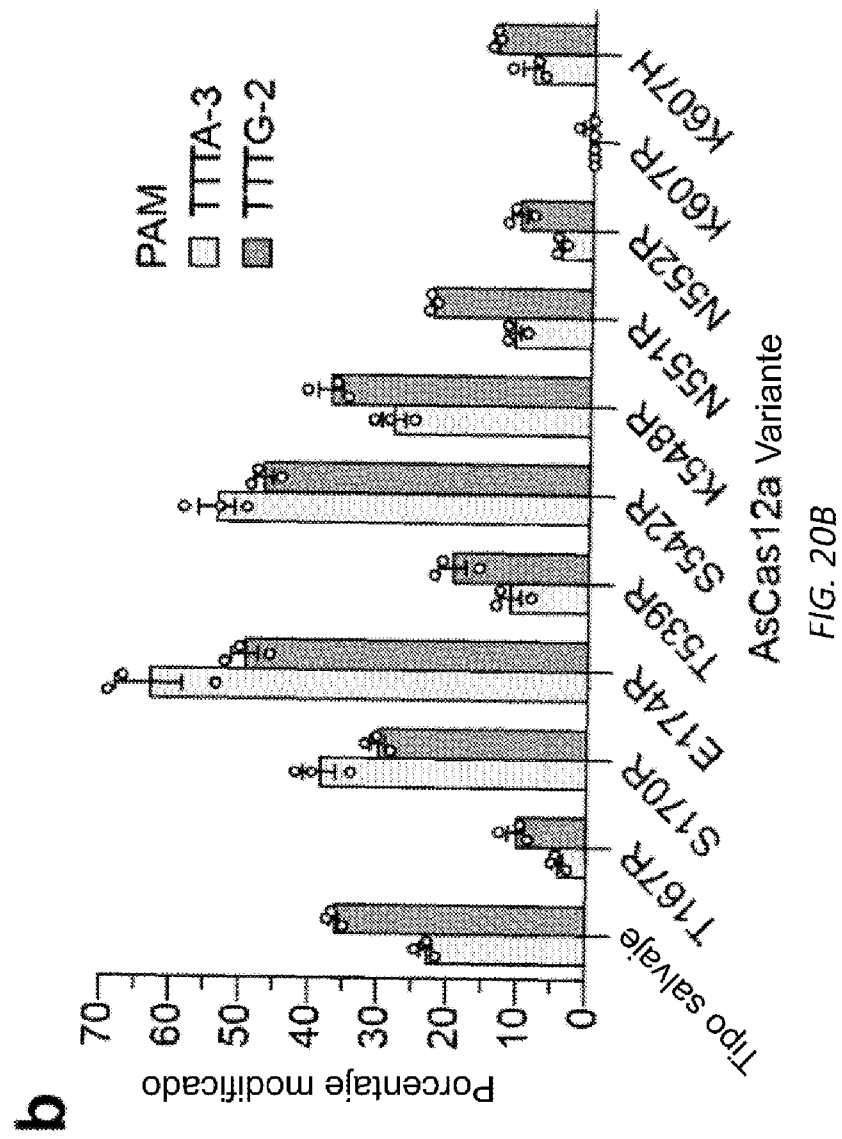
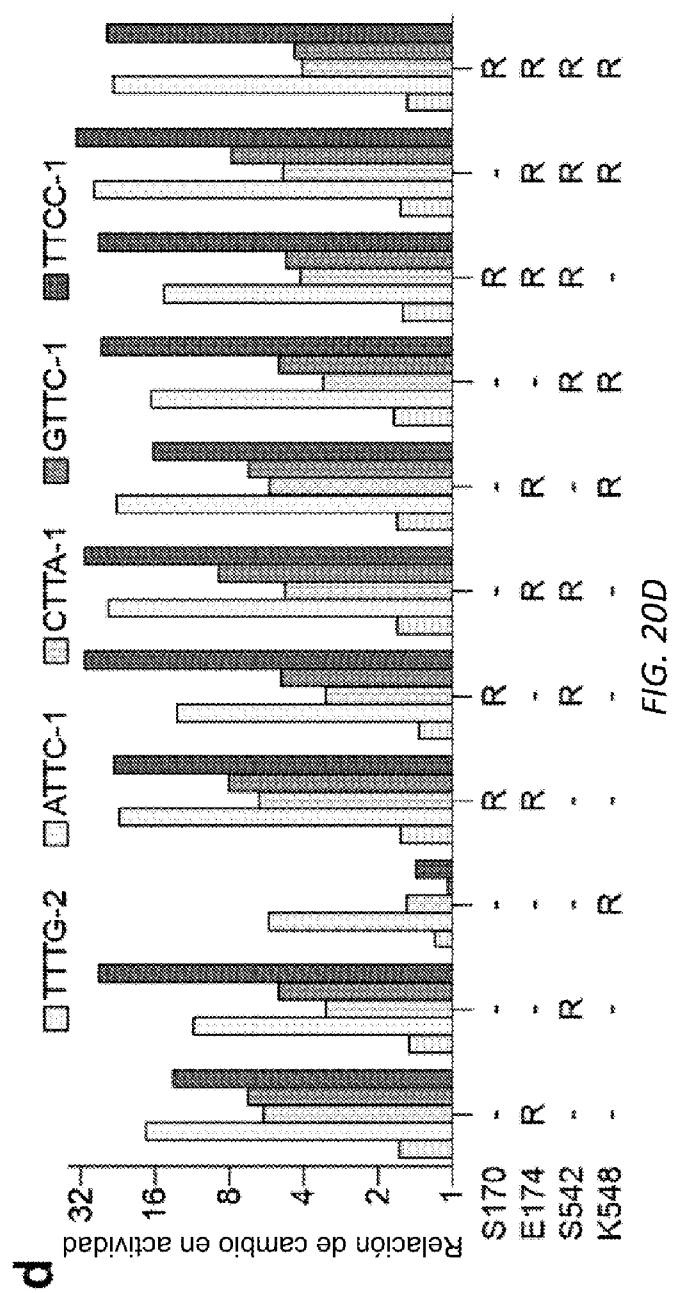
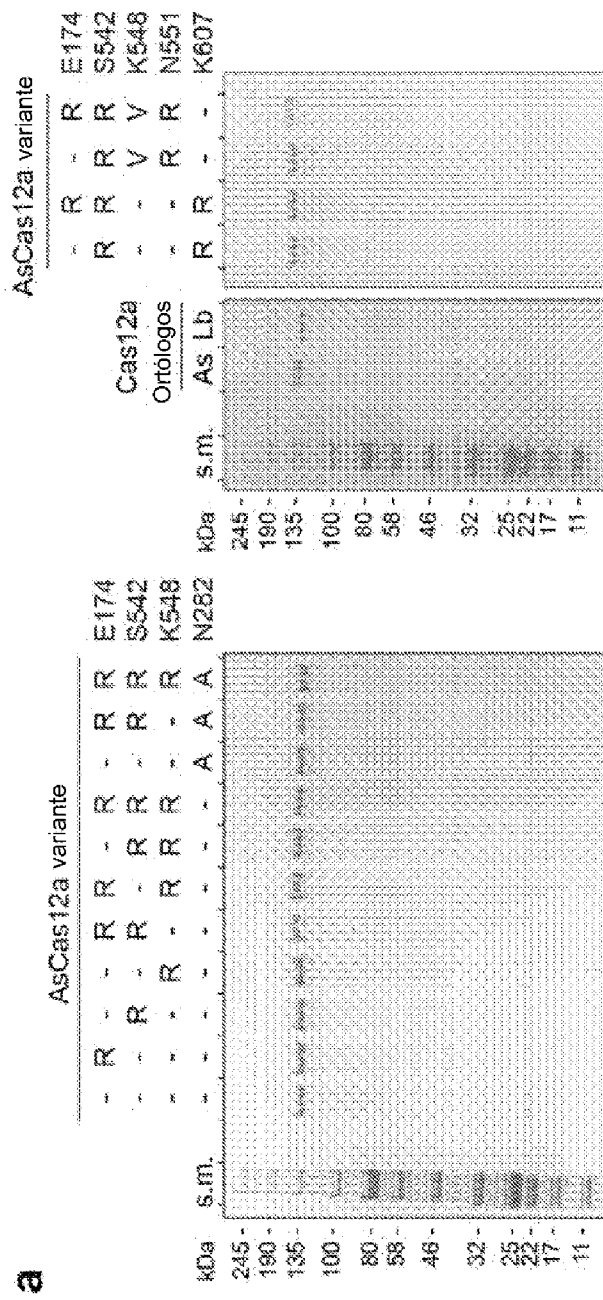


FIG. 20A









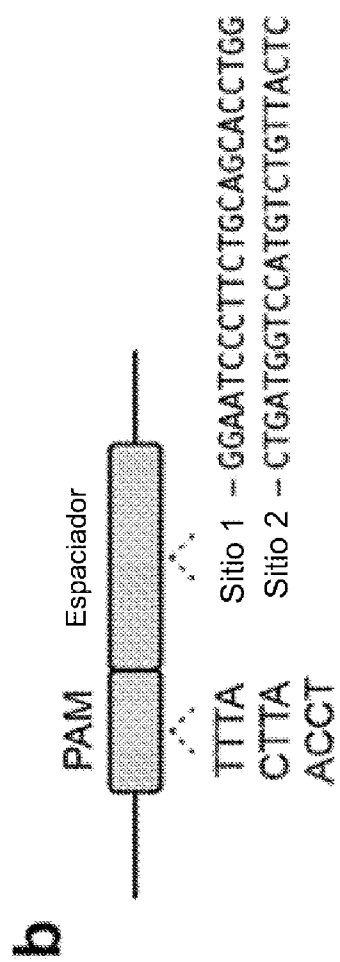


FIG. 21B

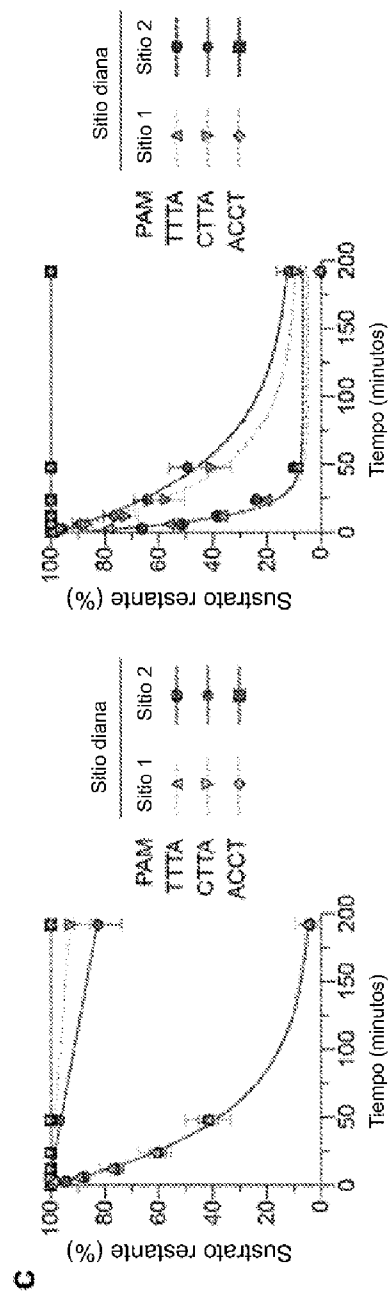


FIG. 21C

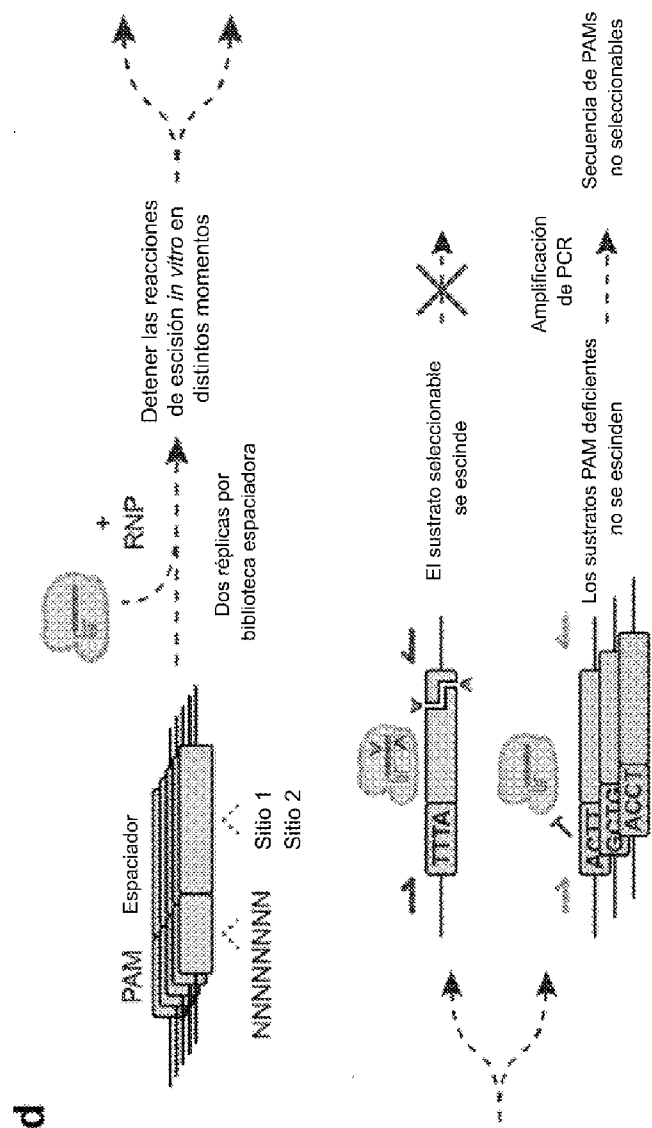


FIG. 21D

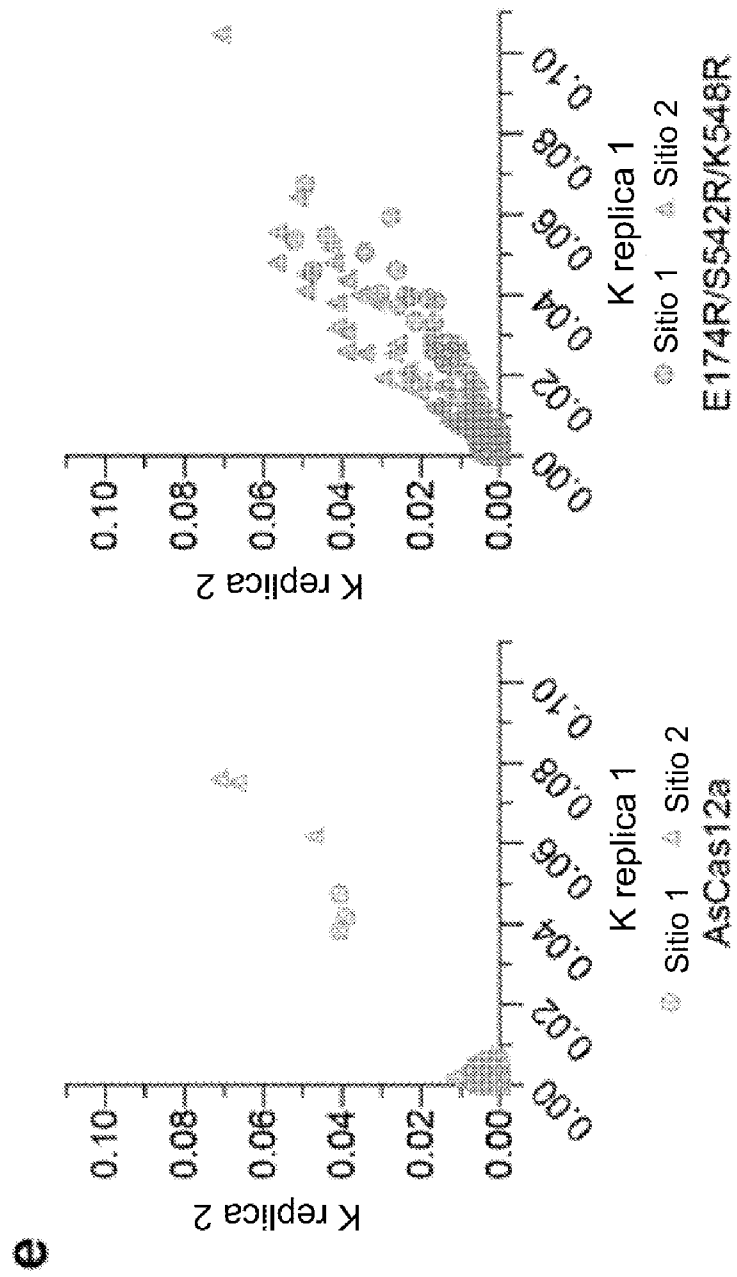
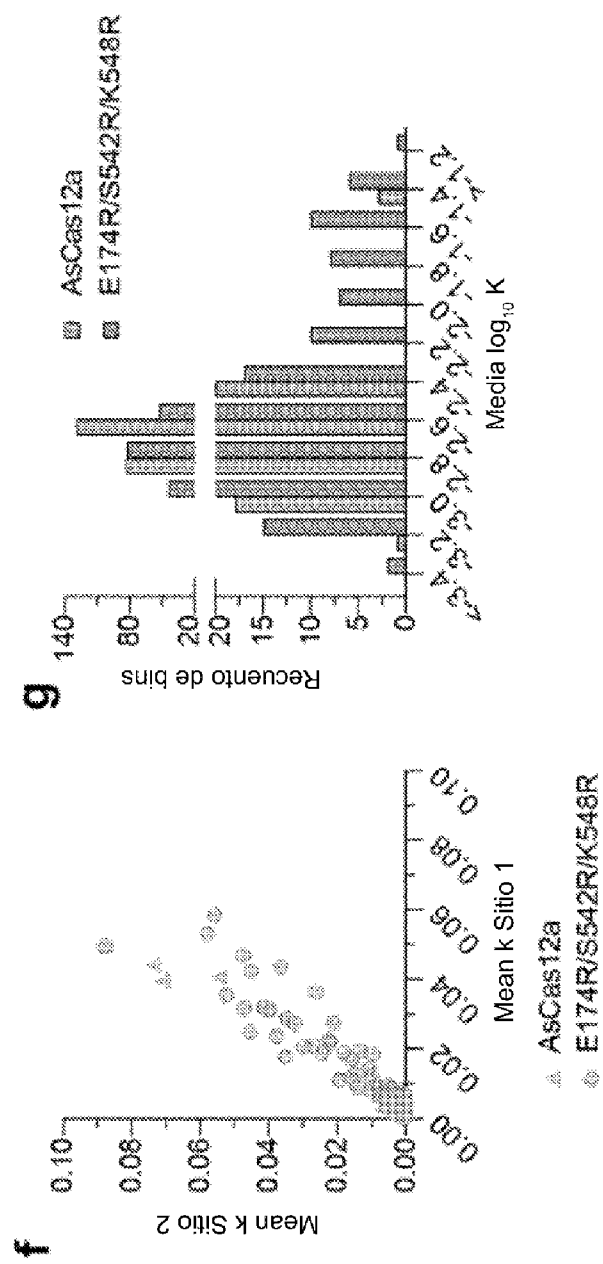


FIG. 21E



FIGs. 21F-G

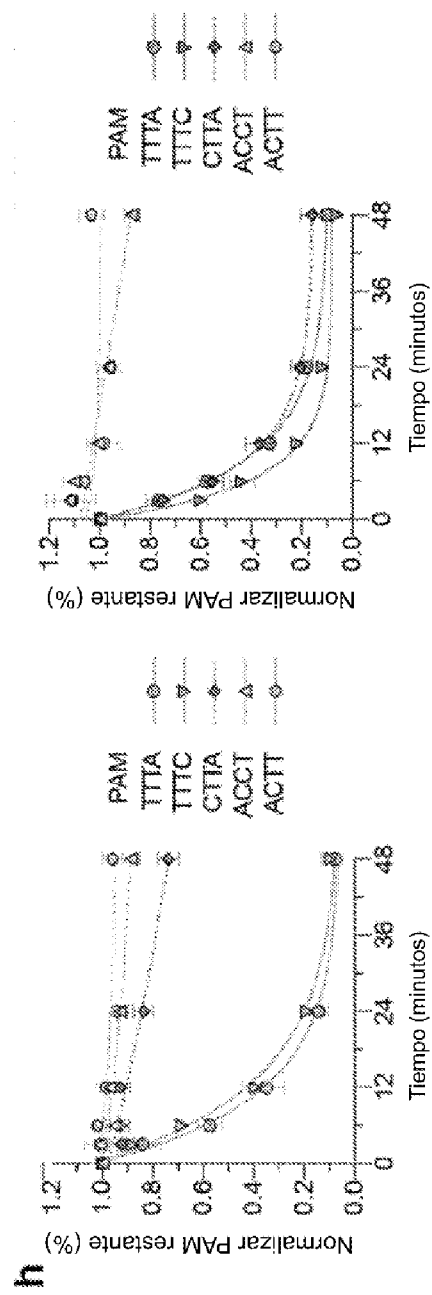


FIG. 21H

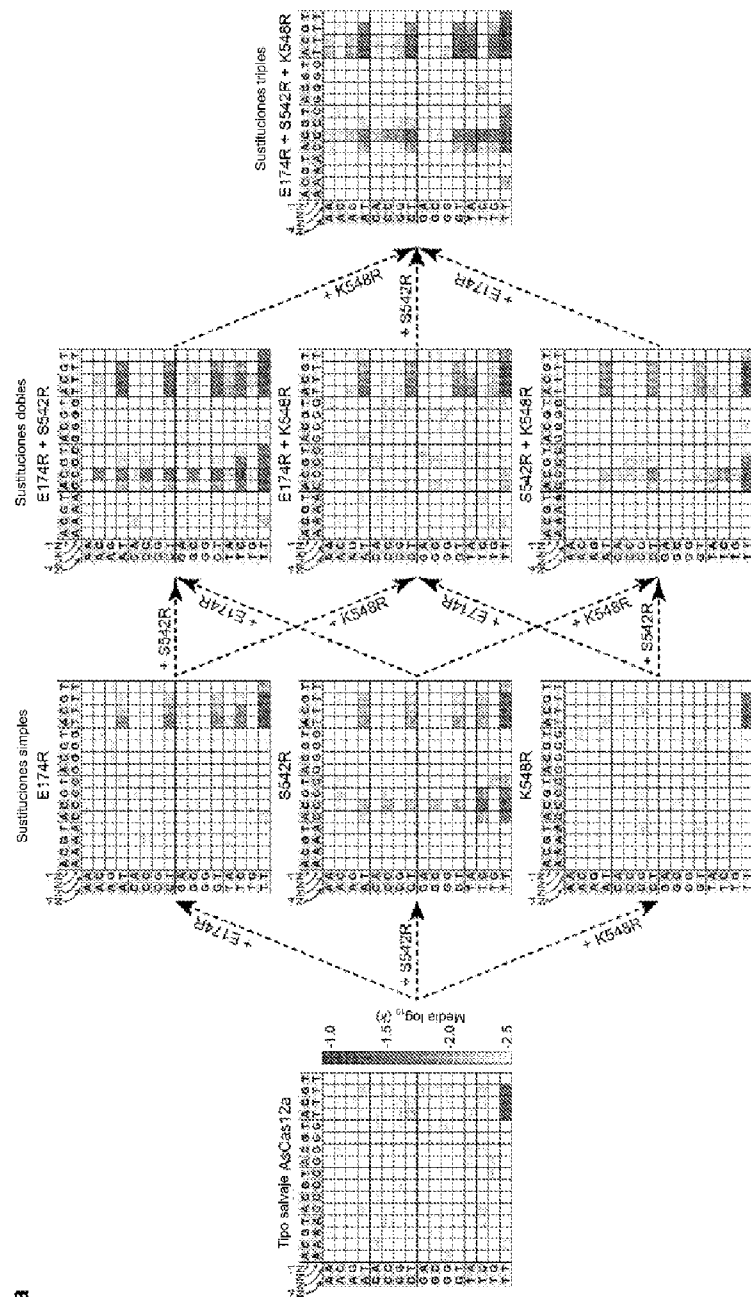


FIG. 22A

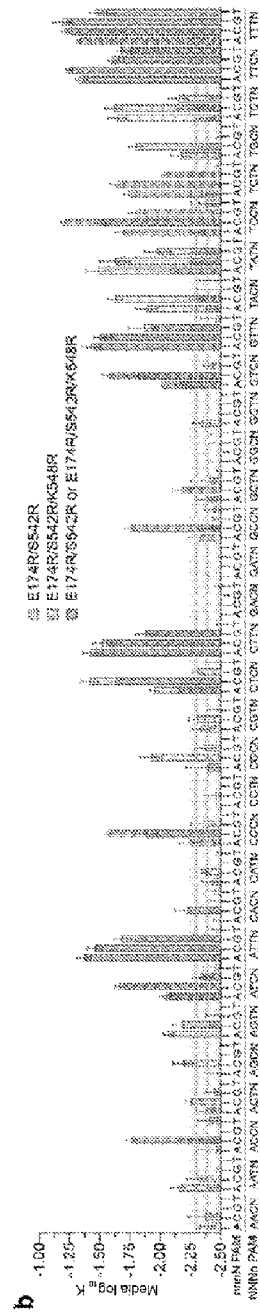


FIG. 22B

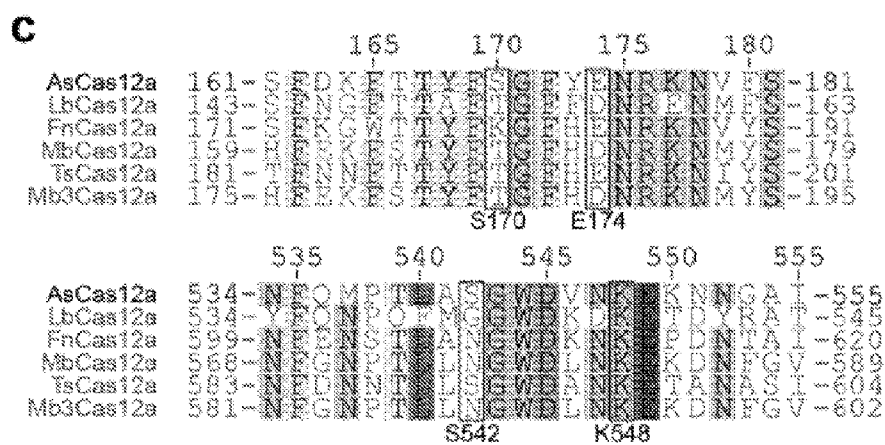


FIG. 22C

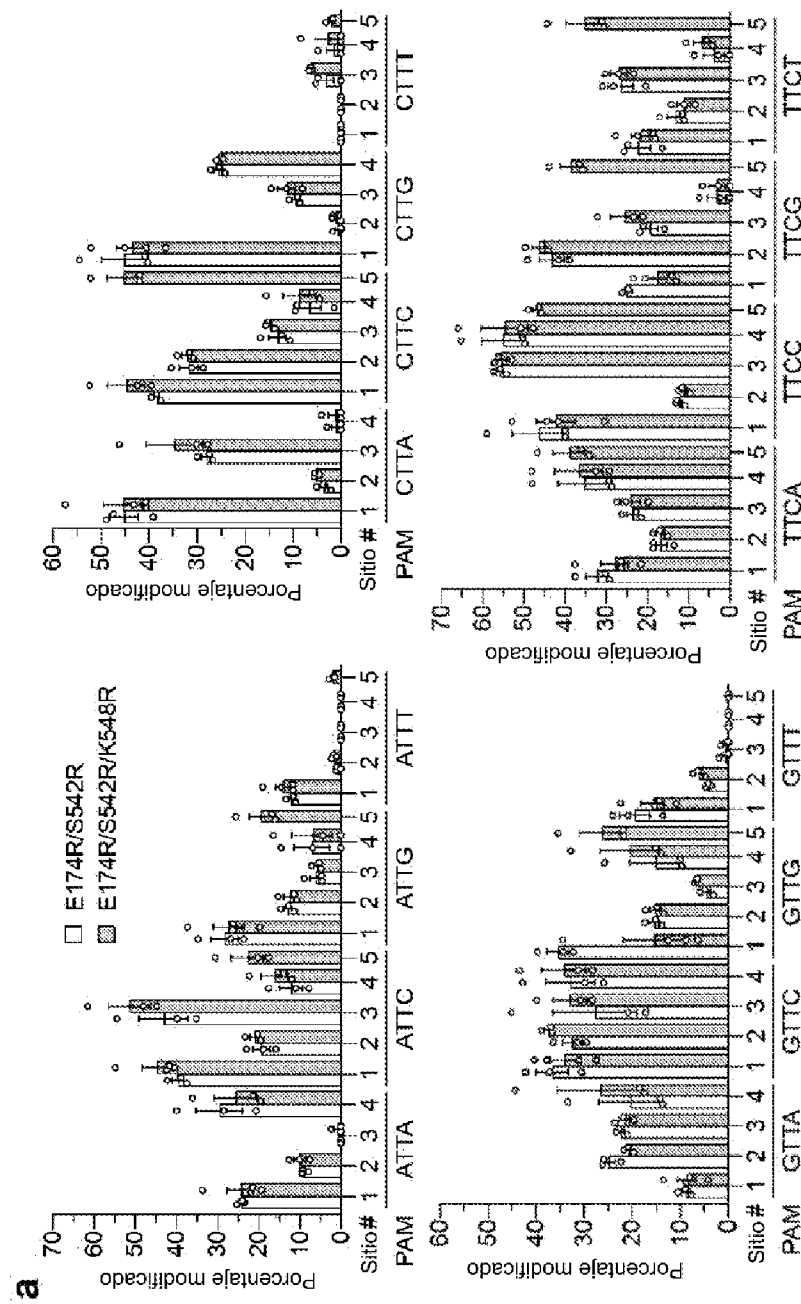
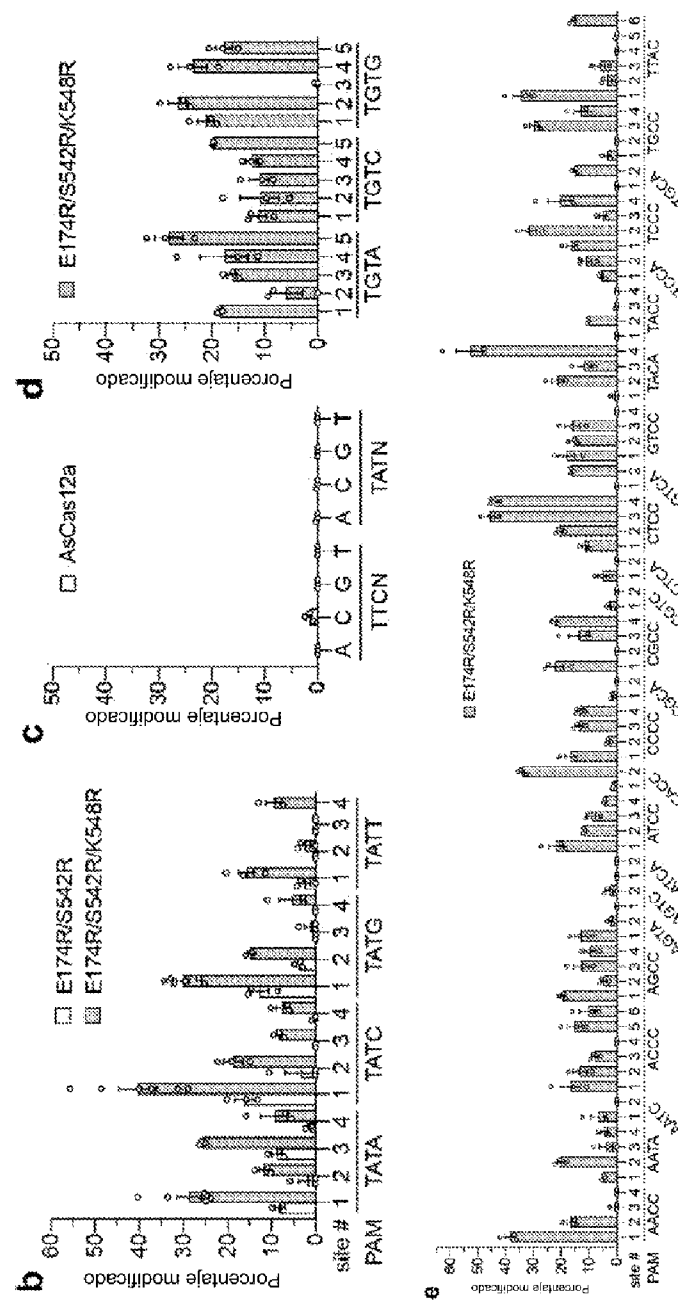


FIG. 23A



FIGS. 23B-E

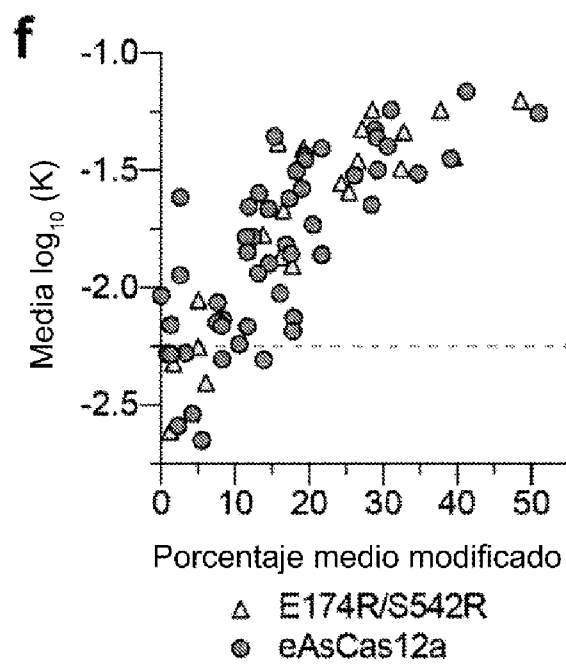
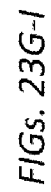


FIG. 23F



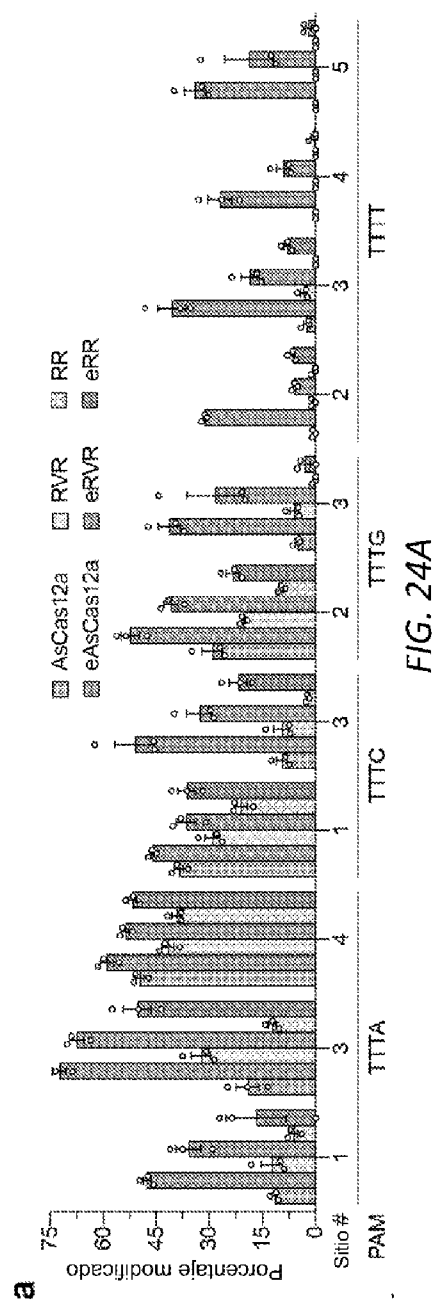


FIG. 24A

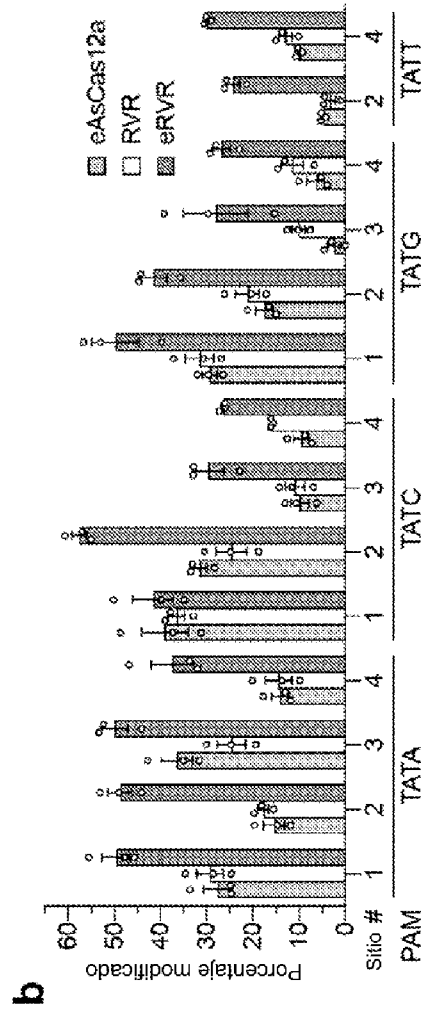
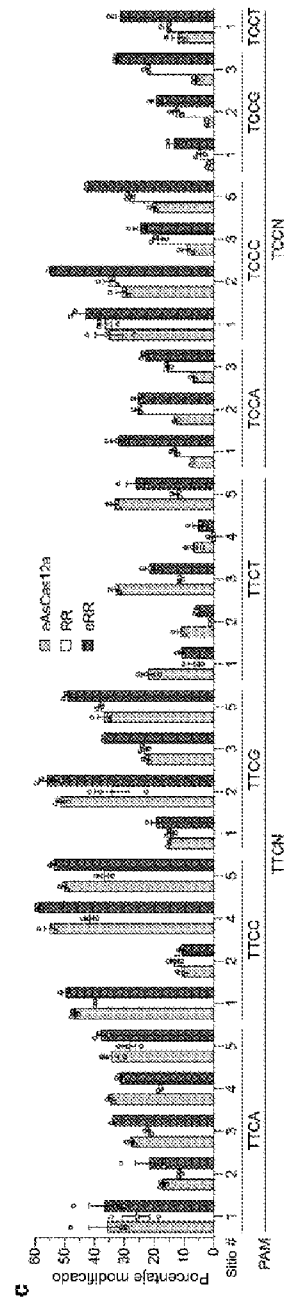


FIG. 24B



FIGs. 24C

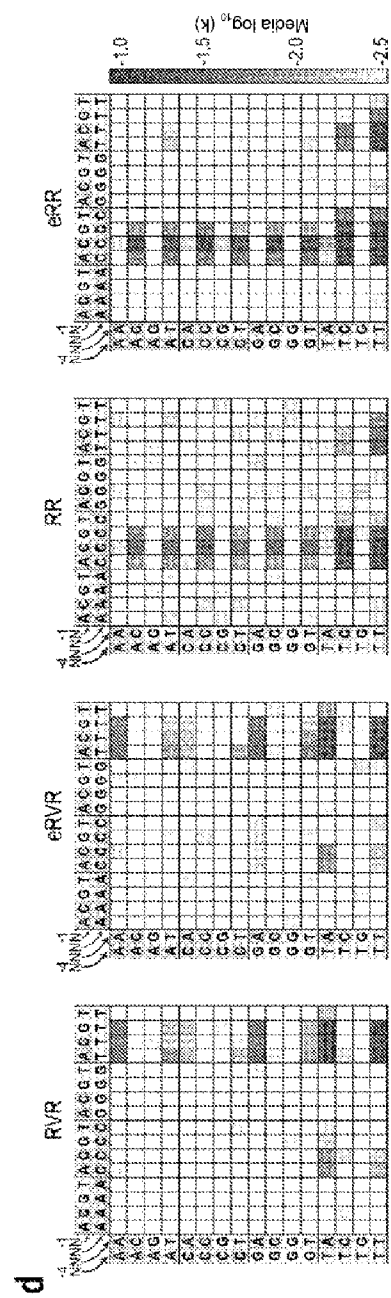


FIG. 24D

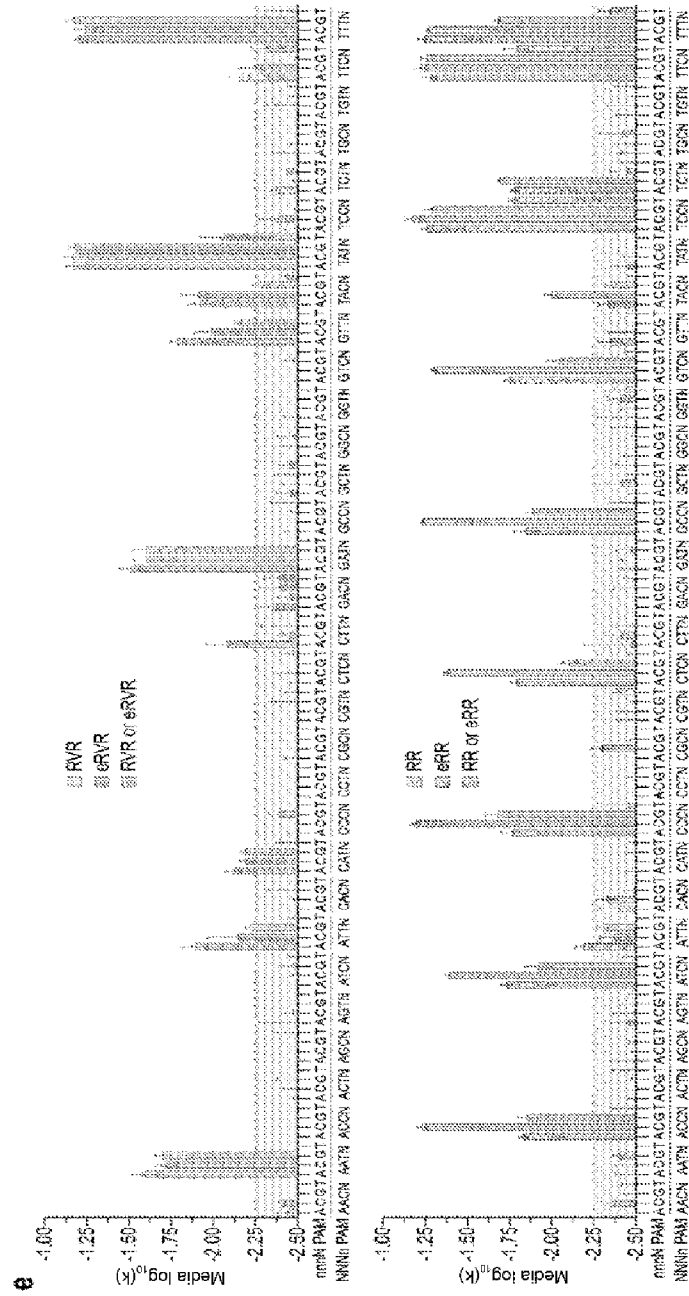
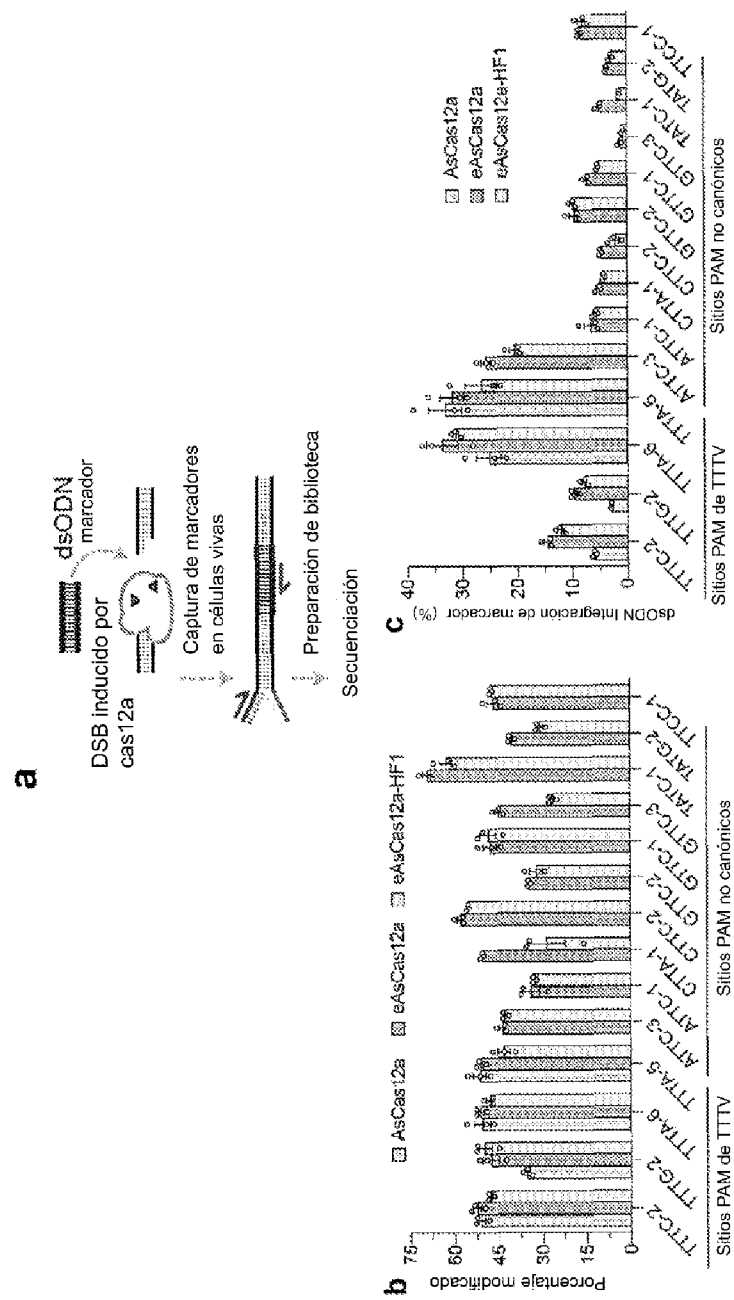
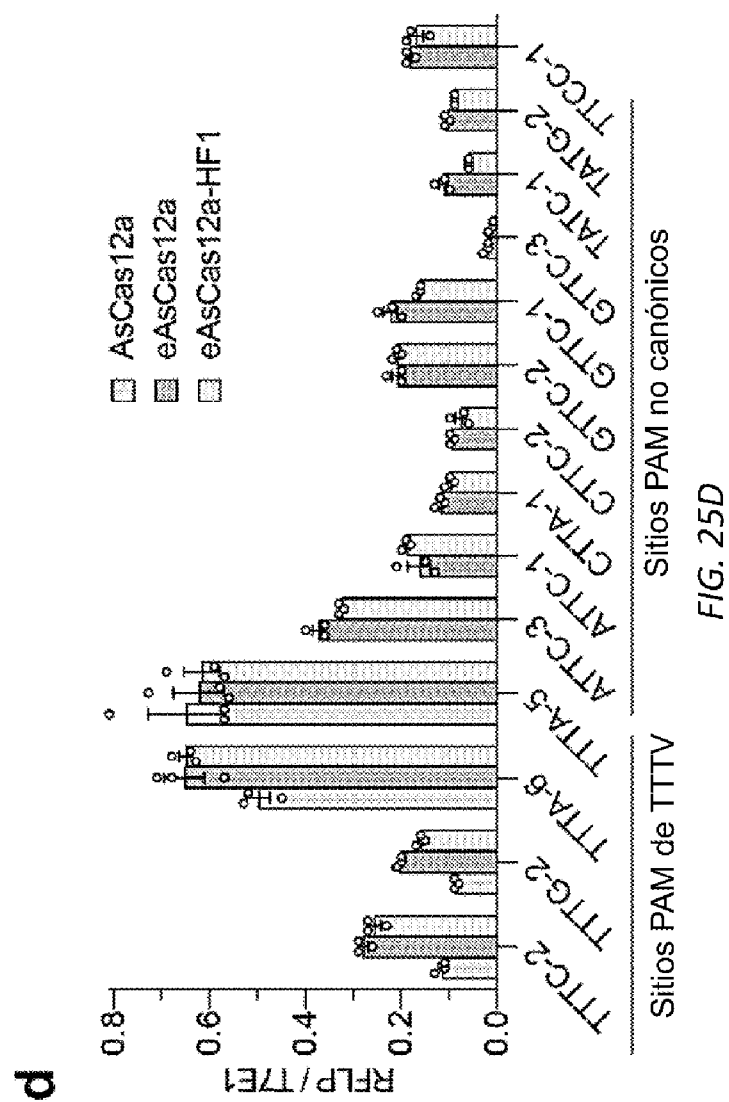
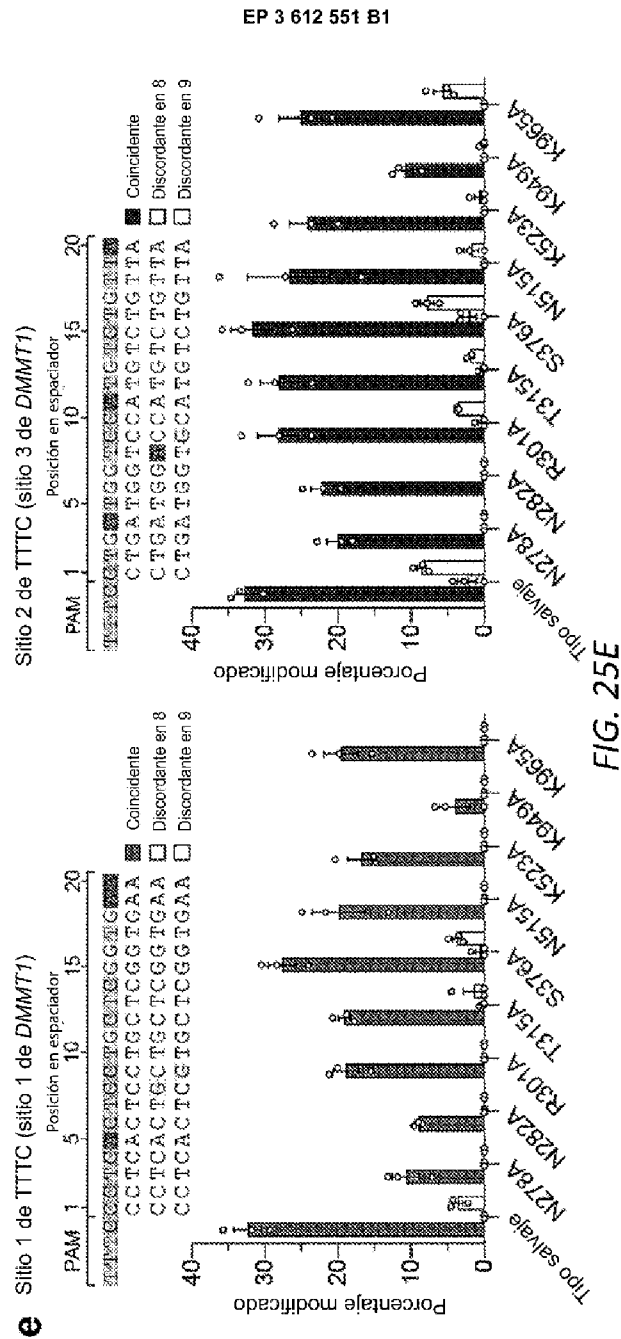


FIG. 24E



FIGS. 25A-C





EP 3 612 551 B1

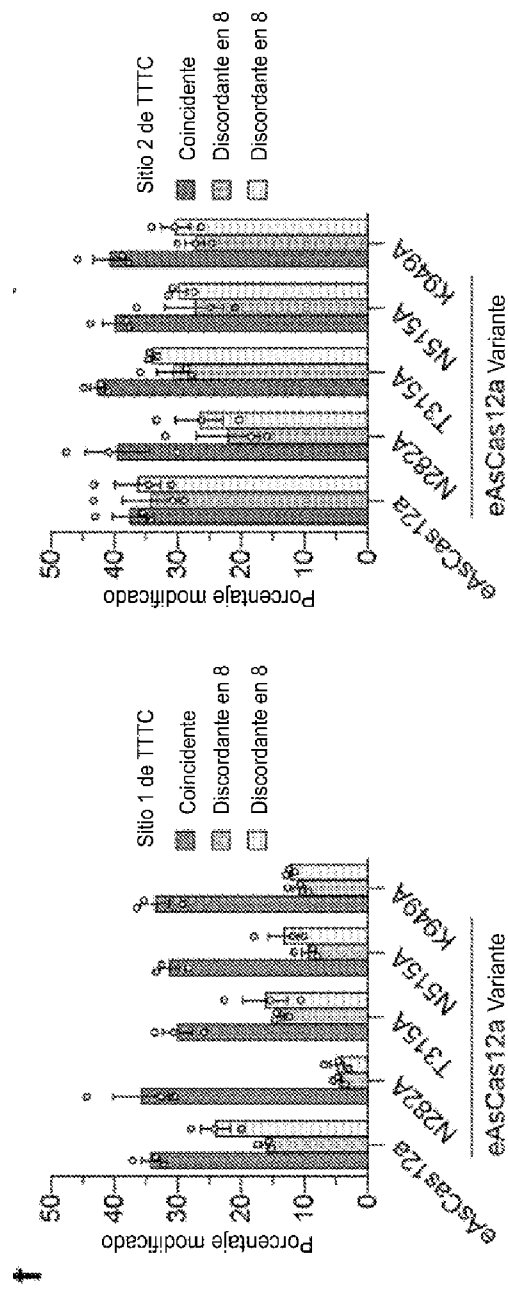


FIG. 25F

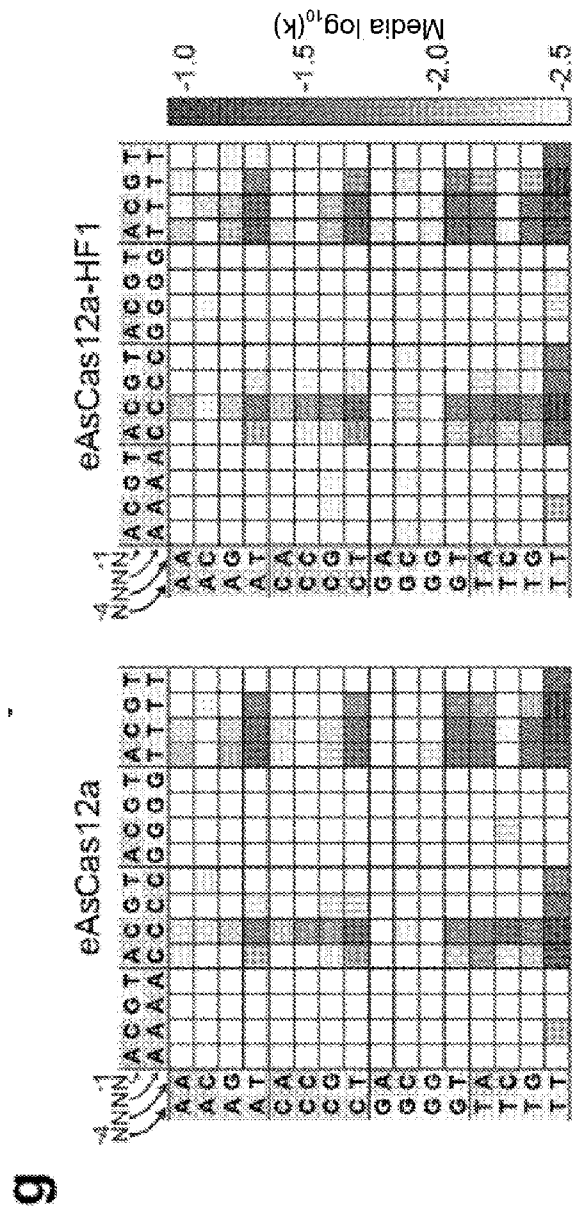


FIG. 25G

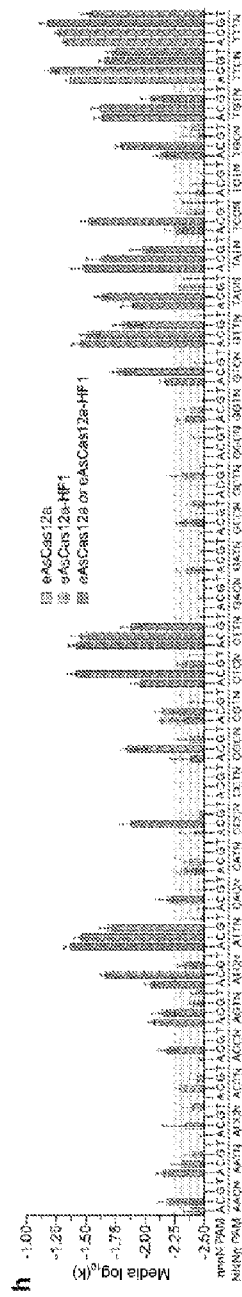


FIG. 25H

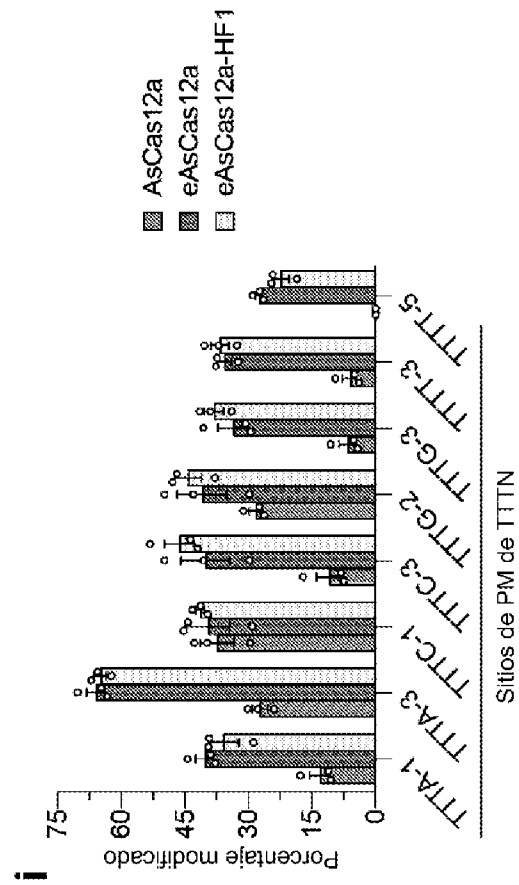
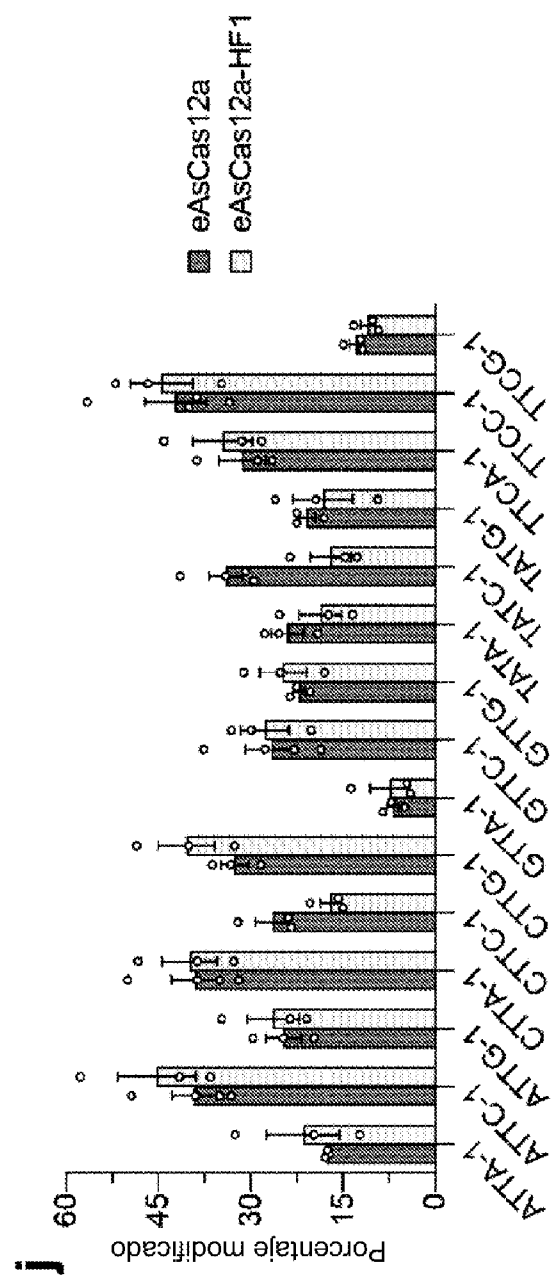


FIG. 251



Sitios PAM no canónicos
 FIG. 25J

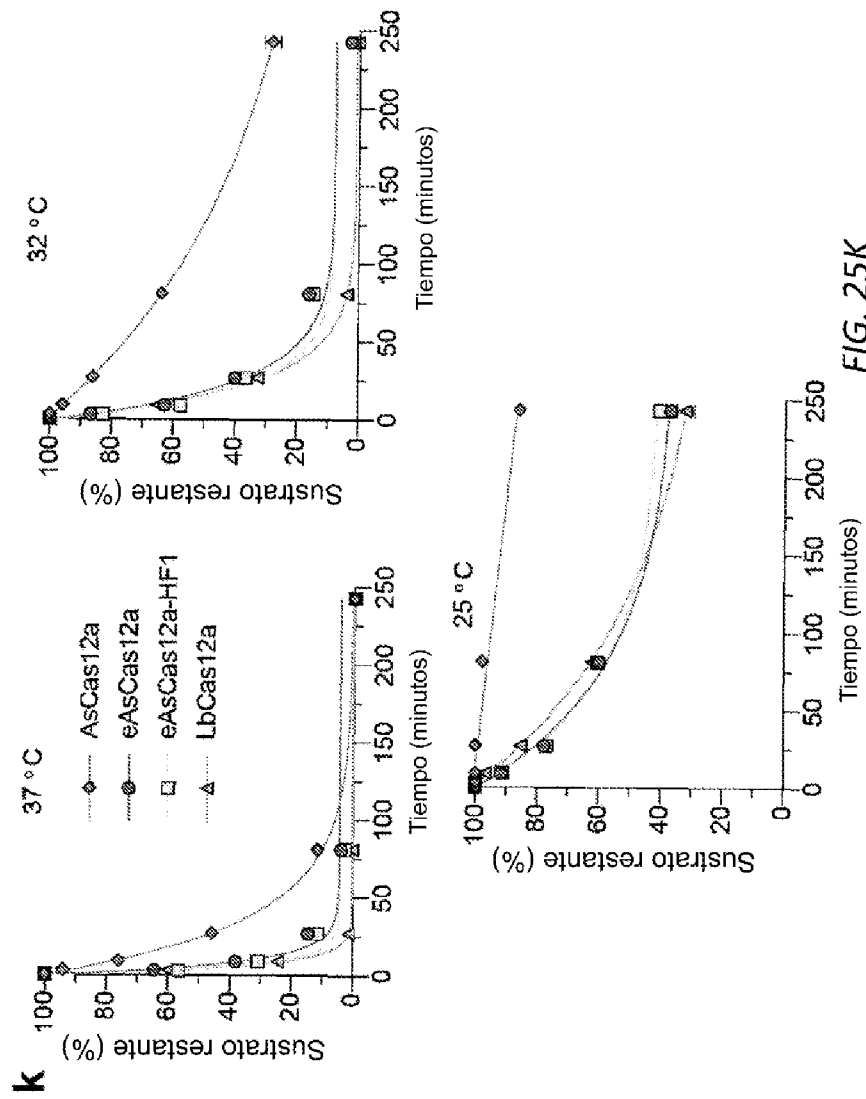


FIG. 25K

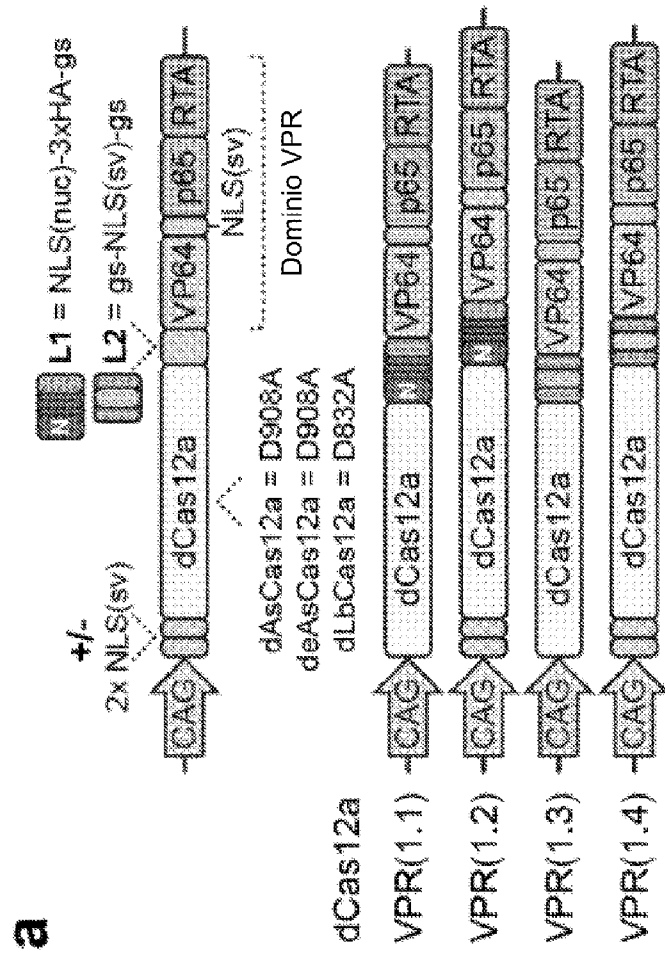
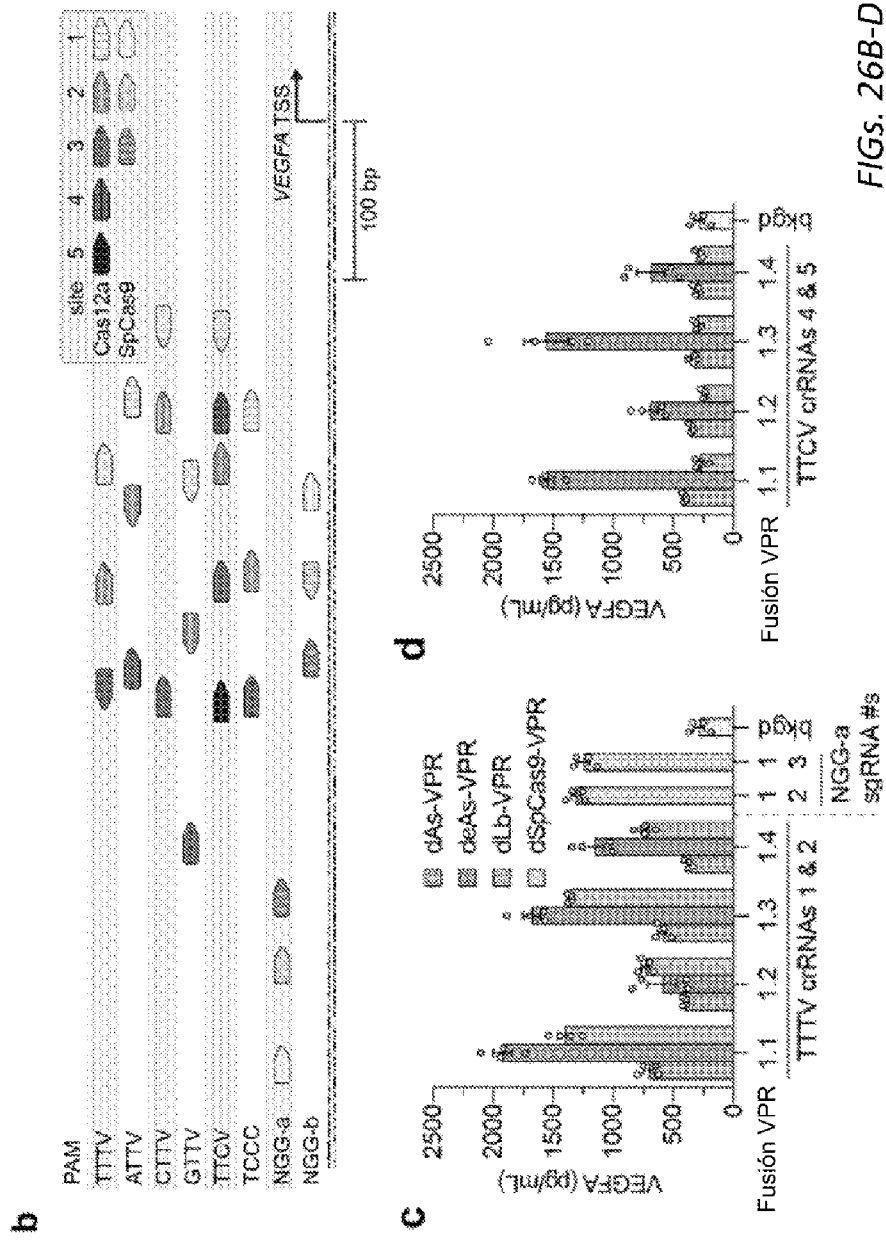
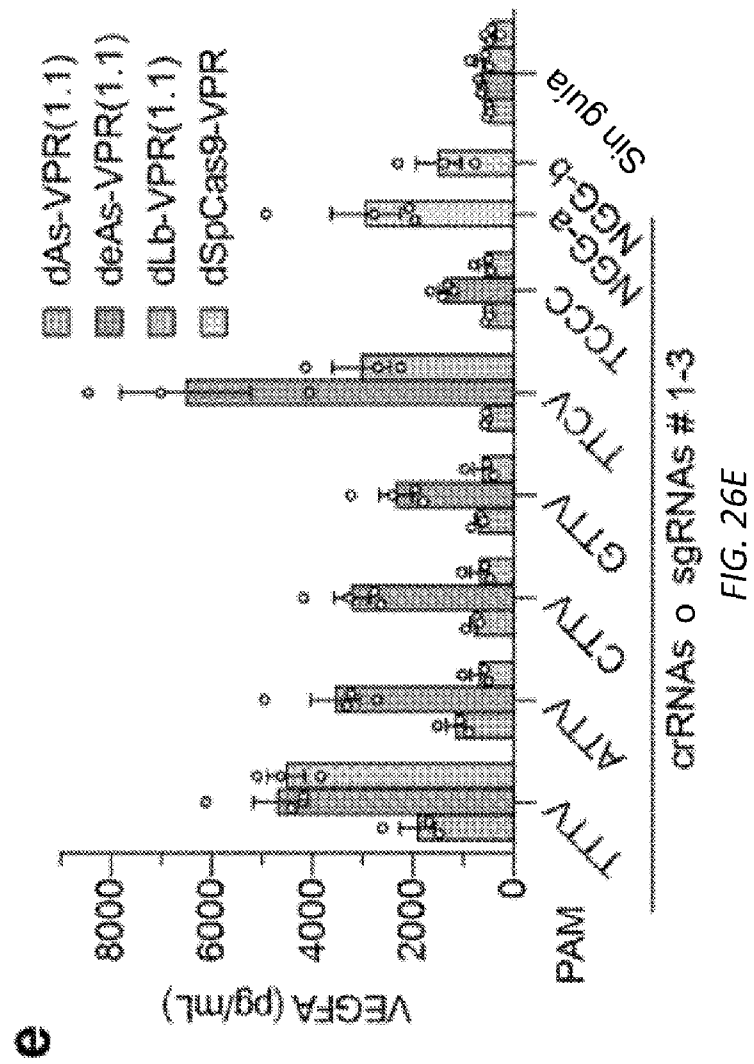


FIG. 26A





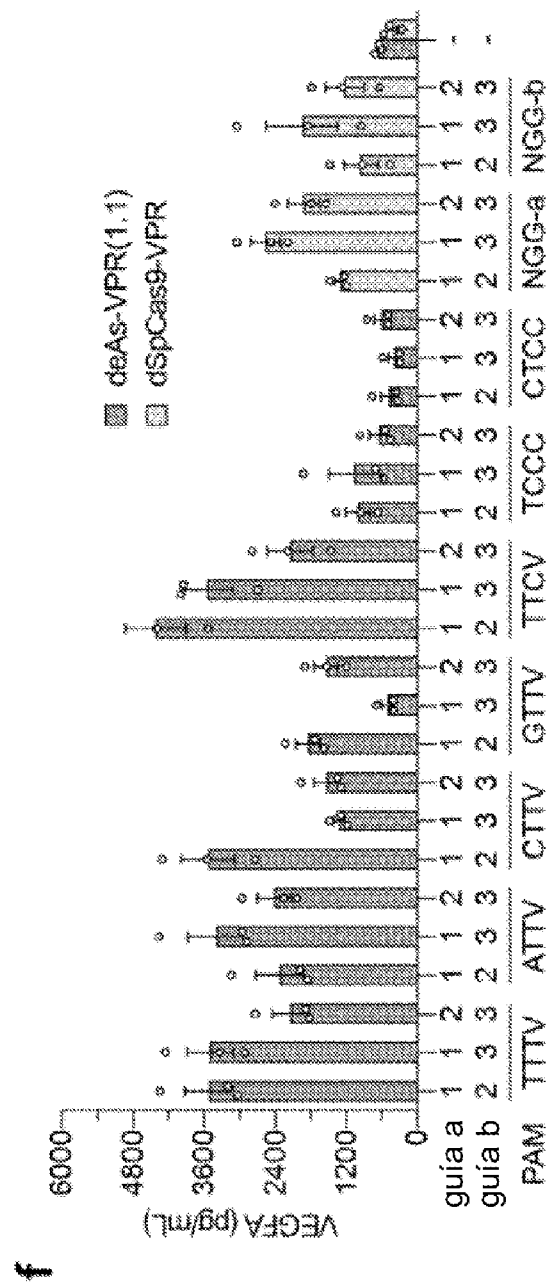


FIG. 26F

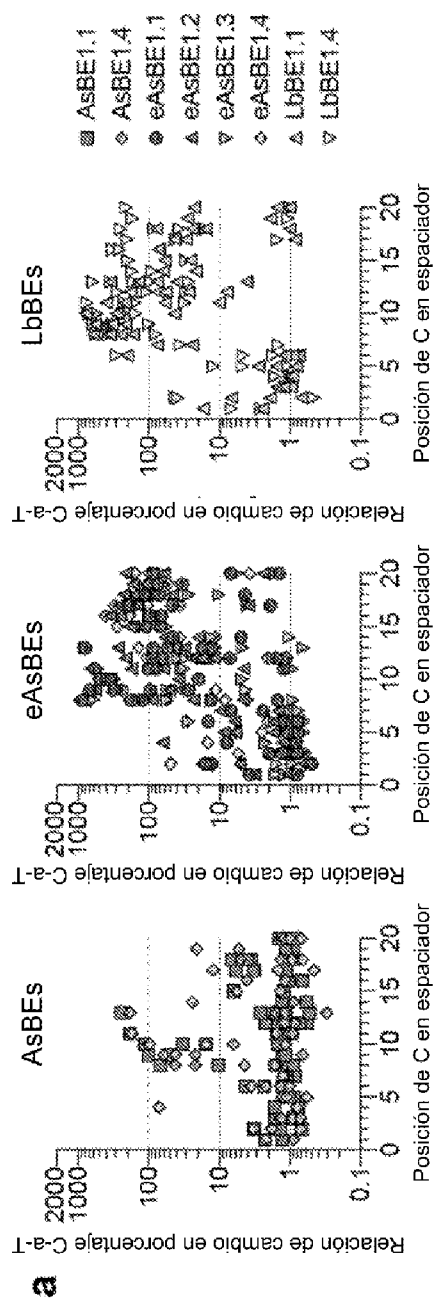


FIG. 27A

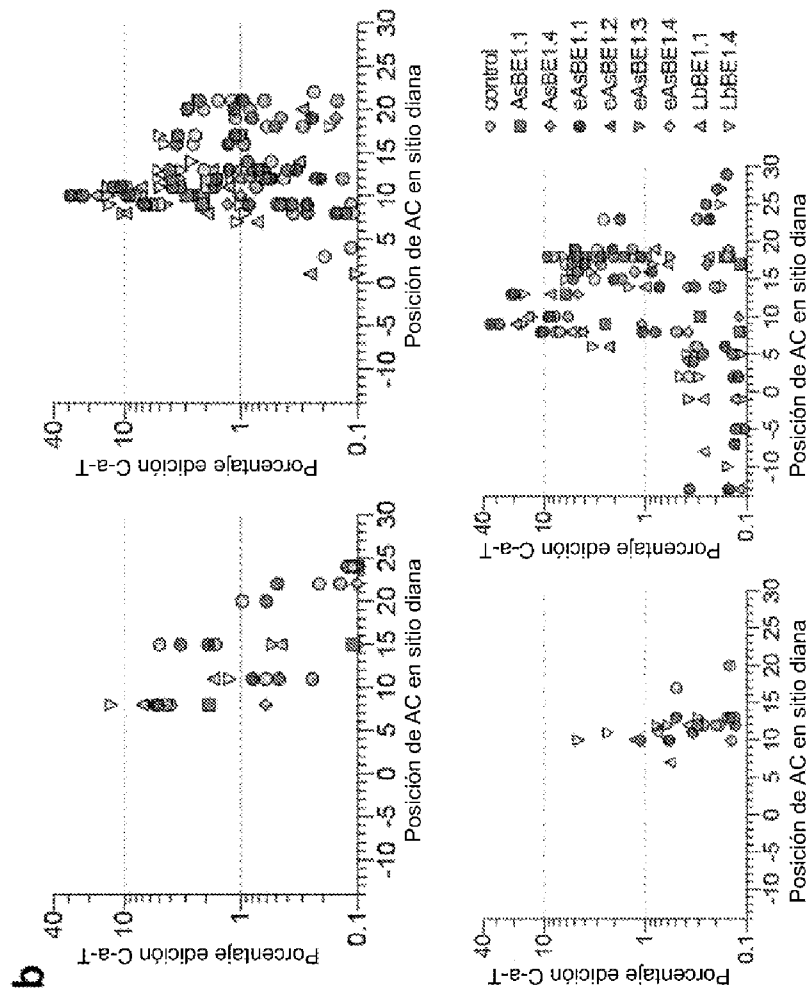


FIG. 27B

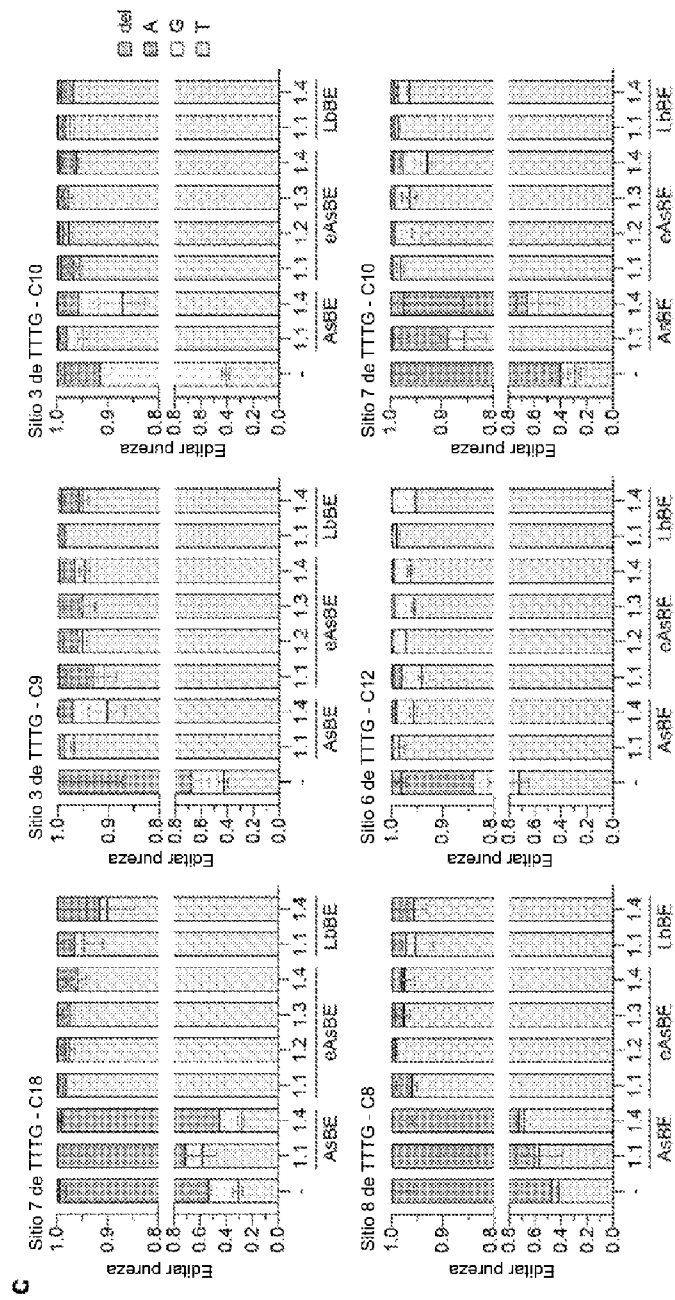


FIG. 27C

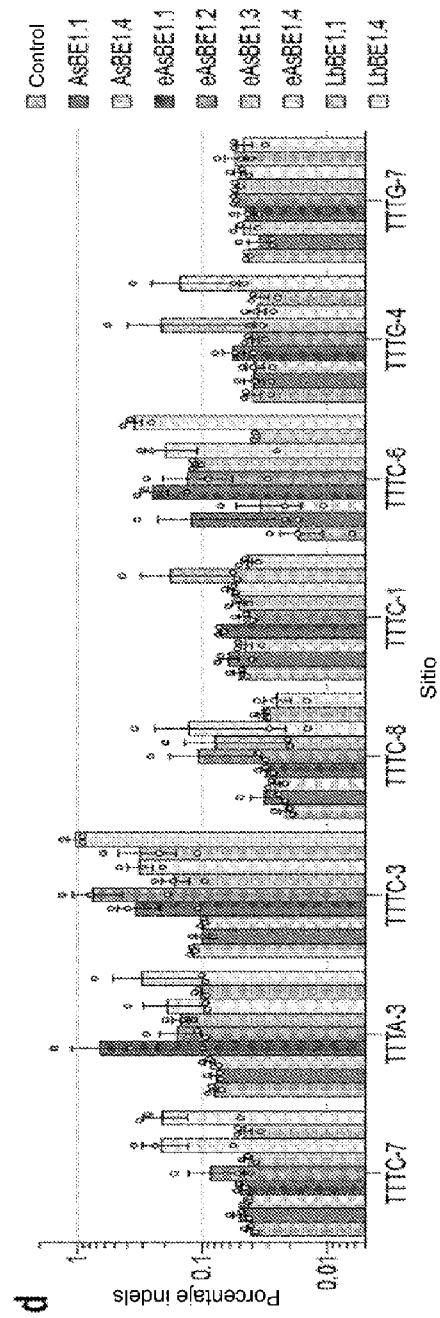
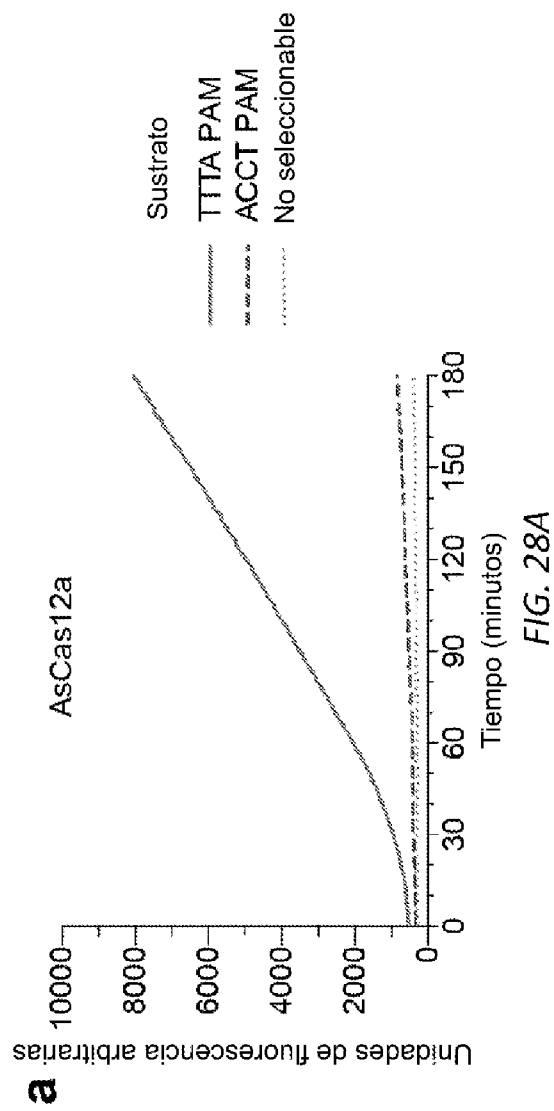


FIG. 27D



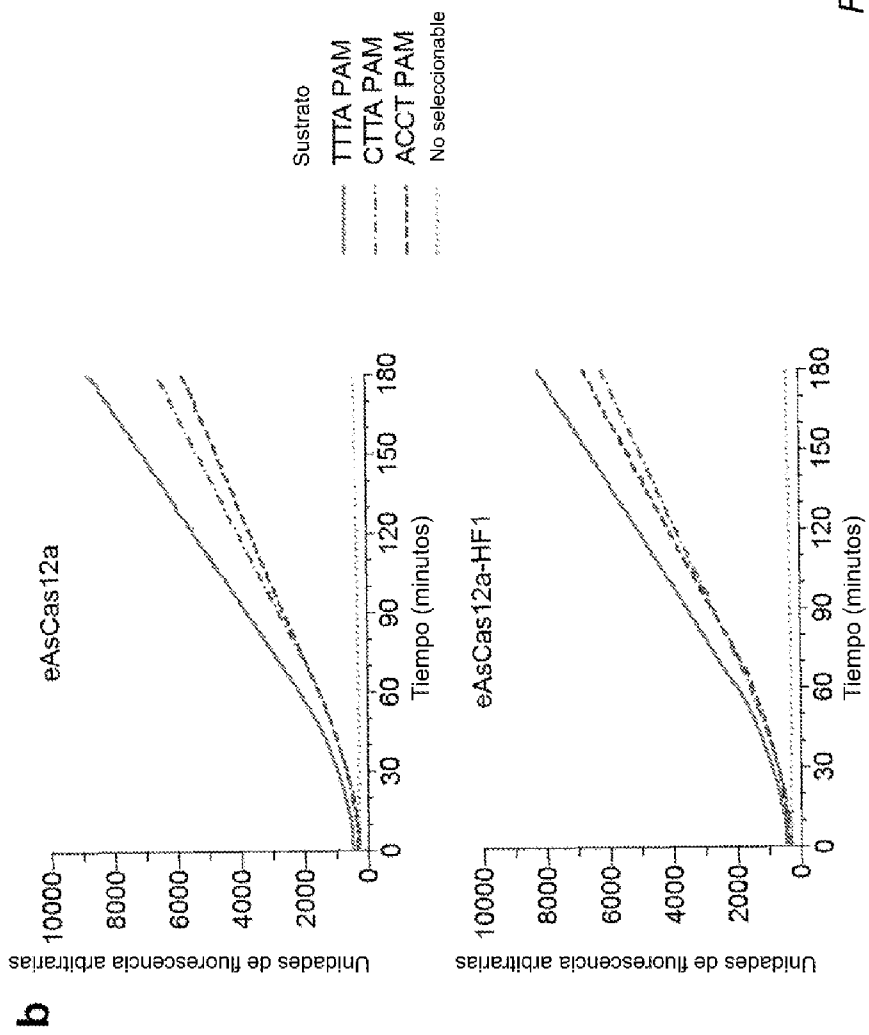


FIG. 28B

REFERENCIAS CITADAS EN LA DESCRIPCIÓN

Esta lista de referencias citadas por el solicitante es sólo a efectos de comodidad del lector. No forma parte del documento de la patente europea. Aunque se ha procurado el máximo cuidado en la compilación de las referencias, no puede excluirse la existencia de errores u omisiones y la OEP declina cualquier responsabilidad al respecto.

Documentos de patente citados en la descripción

- US 62488426 [0001]
- US 62616066 B [0001]
- US 20160208243 A [0054]
- WO 2018022634 A [0069] [0179]
- US 8993233 B [0071] [0073] [0081]
- US 20140186958 A [0071] [0073] [0081]
- US 9023649 B [0071] [0073] [0081]
- WO 2014099744 A [0071] [0073] [0081]
- WO 2014089290 A [0071] [0073] [0081]
- WO 2014144592 A [0071] [0073] [0081]
- WO 144288 A [0071] [0073] [0081]
- WO 2014204578 A [0071] [0073] [0081]
- WO 2014152432 A [0071] [0073] [0081]
- WO 2115099850 A [0071] [0073] [0081]
- US 8697359 B [0071] [0073] [0081]
- US 20160024529 A [0071]
- US 20160024524 A [0071]
- US 20160024523 A [0071]
- US 20160024510 A [0071]
- US 20160017366 A [0071]
- US 20160017301 A [0071]
- US 20150376652 A [0071]
- US 20150356239 A [0071]
- US 20150315576 A [0071]
- US 20150291965 A [0071]
- US 20150252358 A [0071]
- US 20150247150 A [0071]
- US 20150232883 A [0071]
- US 20150232882 A [0071]
- US 20150203872 A [0071]
- US 20150191744 A [0071]
- US 20150184139 A [0071]
- US 20150176064 A [0071]
- US 20150167000 A [0071]
- US 20150166969 A [0071]
- US 20150159175 A [0071]
- US 20150159174 A [0071]
- US 20150093473 A [0071]
- US 20150079681 A [0071]
- US 20150067922 A [0071]
- US 20150056629 A [0071]
- US 20150044772 A [0071]
- US 20150024500 A [0071]
- US 20150024499 A [0071]
- US 20150020223 A [0071]
- US 20140356867 A [0071]
- US 20140295557 A [0071]
- US 20140273235 A [0071]
- US 20140273226 A [0071]
- US 20140273037 A [0071]
- US 20140189896 A [0071]
- US 20140113376 A [0071]
- US 20140093941 A [0071]
- US 20130330778 A [0071]
- US 20130288251 A [0071]
- US 20120088676 A [0071]
- US 20110300538 A [0071]
- US 20110236530 A [0071]
- US 20110217739 A [0071]
- US 20110002889 A [0071]
- US 20100076057 A [0071] [0073] [0081]
- US 20110189776 A [0071] [0073] [0081]
- US 20110223638 A [0071] [0073] [0081]
- US 20130130248 A [0071] [0073] [0081]
- US 20150050699 A [0071] [0073] [0081]
- US 20150071899 A [0071] [0073] [0081]
- US 20150045546 A [0071]
- US 20150031134 A [0071]
- US 20140377868 A [0071]
- US 20140357530 A [0071]
- US 20140349400 A [0071]
- US 20140335620 A [0071]
- US 20140335063 A [0071]
- US 20140315985 A [0071]
- US 20140310830 A [0071]
- US 20140310828 A [0071]
- US 20140309487 A [0071]
- US 20140304853 A [0071]
- US 20140298547 A [0071]
- US 20140295556 A [0071]
- US 20140294773 A [0071]
- US 20140287938 A [0071]
- US 20140273234 A [0071]
- US 20140273232 A [0071]
- US 20140273231 A [0071]
- US 20140273230 A [0071]
- US 20140271987 A [0071]
- US 20140256046 A [0071]

- US 20140248702 A [0071]
- US 20140242702 A [0071]
- US 20140242700 A [0071]
- US 20140242699 A [0071]
- US 20140242664 A [0071]
- US 20140234972 A [0071]
- US 20140227787 A [0071]
- US 20140212869 A [0071]
- US 20140201857 A [0071]
- US 20140199767 A [0071]
- US 20140186919 A [0071]
- US 20140186843 A [0071]
- US 20140179770 A [0071]
- US 20140179006 A [0071]
- US 20140170753 A [0071]
- WO 2008108989 A [0071] [0073] [0081]
- WO 2010054108 A [0071] [0073] [0081]
- WO 2012164565 A [0071] [0073] [0081]
- WO 2013098244 A [0071] [0073] [0081]
- WO 2013176772 A [0071] [0073] [0081]
- US 61652086 [0071]
- US 20170362644 A [0072] [0103]
- WO 2017219027 A1 [0072] [0103]
- WO 2014124284 A [0073]

Bibliografía no relacionada con patentes citada en la descripción

- ZETSCHE et al. *Cell*, 2015, vol. 163, 759-771 [0004] [0060]
- SHMAKOV et al. *Mol Cell*, 05 November 2015, vol. 60 (3), 385-97 [0004]
- KLEINSTIVER et al. *Nat Biotechnol.*, August 2016, vol. 34 (8), 869-74 [0004]
- KIM et al. *Nat Biotechnol.*, August 2016, vol. 34 (8), 863-8 [0004]
- KLEINSTIVER et al. *Nat Biotechnol.*, 2016, vol. 34, 869-74 [0050] [0179]
- CHAVEZ et al. *Nat Methods.*, 2015, vol. 12, 326-8 [0050] [0183]
- YAMANO et al. *Cell*, 05 May 2016, vol. 165 (4), 949-62 [0050] [0067] [0157]
- WRIGHT et al. *Cell*, 2016, vol. 164, 29-44 [0053]
- SMITH, T. F. ; M. S. WATERMAN. *J Mol Biol*, 1981, vol. 147, 195-7 [0063]
- SMITH; WATERMAN. *Advances in Applied Mathematics*, 1981, 482-489 [0063]
- SCHWARZ ; DAYHOF. *Atlas of Protein Sequence and Structure*. 1979, 353-358 [0063]
- ALTSCHUL, S. F. ; W. GISH et al. *Basic Local Alignment Search Tool. J Mol Biol*, 1990, vol. 215, 403-10 [0063]
- FONFARA et al. *Nature*, 28 April 2016, vol. 532 (7600), 517-21 [0067]
- DONG et al. *Nature*, 28 April 2016, vol. 532 (7600), 522-6 [0067]
- ZETSCHE et al. *Cell*, 22 October 2015, vol. 163 (3), 759-71 [0067]
- MAKAROVA et al. *Evolution and classification of the*
- CRISPR-Cas systems. *Nature Reviews Microbiology*, June 2011, vol. 9 (6), 467-477 [0071]
- WIEDENHEFT et al. *RNA-guided genetic silencing systems in bacteria and archaea. Nature*, 16 February 2012, vol. 482, 331-338 [0071]
- GASIUNAS et al. *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences USA*, 04 September 2012, vol. 109 (39), E2579-E2586 [0071]
- JINEK et al. *A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science*, 17 August 2012, vol. 337, 816-821 [0071]
- CARROLL. *A CRISPR Approach to Gene Targeting. Molecular Therapy*, September 2012, vol. 20 (9), 1658-1660 [0071]
- AL-ATTAR et al. *Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs): The Hallmark of an Ingenious Antiviral Defense Mechanism in Prokaryotes. Biol Chem.*, 2011, vol. 392 (4), 277-289 [0071]
- HALE et al. *Essential Features and Rational Design of CRISPR RNAs That Function With the Cas RAMP Module Complex to Cleave RNAs. Molecular Cell*, 2012, vol. 45 (3), 292-302 [0071]
- EAST-SELETSKY et al. *Nature*, 13 October 2016, vol. 538 (7624), 270-273 [0072] [0103]
- GOOTENBERG et al. *Science*, 28 April 2017, vol. 356 (6336), 438-442 [0072] [0103]
- GOOTENBERG et al. *Science*, 2018 [0072] [0188]
- CHEN et al. *Science*, 15 February 2018 [0072]
- *Science*, 15 February 2018 [0072]
- SADOWSKI et al. *Nature*, 1988, vol. 335, 563-564 [0073]

- RUBEN et al. Science, 1991, vol. 251, 1490-93 [0073]
- CHAVEZ et al. Nat Methods., April 2015, vol. 12 (4), 326-8 [0073]
- BEERLI et al. PNAS USA, 1998, vol. 95, 14628-14633 [0073]
- IYER et al. Cell Cycle., 01 June 2009, vol. 8 (11), 1698-710 [0075]
- YANG et al. J Genet Genomics, 20 September 2017, vol. 44 (9), 423-437 [0077]
- SAVVA et al. Genome Biol., 28 December 2012, vol. 13 (12), 252 [0078]
- KEEGAN et al. RNA., September 2017, vol. 23 (9), 1317-1328 [0078]
- SCHAUB ; KELLER. Biochimie., August 2002, vol. 84 (8), 791-803 [0078]
- GAUDELLI et al. Nature, 23 November 2017, vol. 551 (7681), 464-471 [0078] [0186]
- WOLF et al. EMBO J., 15 July 2002, vol. 21 (14), 3841-51 [0078]
- MOL et al. Cell, 1995, vol. 82, 701-708 [0079]
- KOMOR et al. Nature, 19 May 2016, vol. 533 (7603) [0079]
- KOMOR et al. Sci Adv., 30 August 2017, vol. 3 (8), eaao4774 [0079] [0186]
- KERYER-BIBENS et al. Biol. Cell, 2008, vol. 100, 125-138 [0081]
- CARON et al. Mol Ther., 2001, vol. 3 (3), 310-8 [0083]
- LANGE. Cell-Penetrating Peptides: Processes and Applications. CRC Press, 2002 [0083]
- EL-ANDALOUSSI et al. Curr Pharm Des., 2005, vol. 11 (28), 3597-611 [0083]
- DESHAYES et al. Cell Mol Life Sci., 2005, vol. 62 (16), 1839-49 [0083]
- FRANKEL et al. Cell, 1988, vol. 55, 1189-1193 [0084]
- VIVES et al. J. Biol. Chem., 1997, vol. 272, 16010-16017 [0084]
- DEROSI et al. J. Biol. Chem., 1994, vol. 269, 10444-10450 [0084]
- WENDER et al. Proc. Natl. Acad. Sci. USA, 2000, vol. 97, 13003-13008 [0084]
- FUTAKI et al. J. Biol. Chem., 2001, vol. 276, 5836-5840 [0084]
- POOGA et al. Nat. Biotechnol., 1998, vol. 16, 857-861 [0084]
- STETSENKO et al. J. Org. Chem., 2000, vol. 65, 4900-4909 [0085]
- GAIT et al. Cell. Mol. Life. Sci., 2003, vol. 60, 844-853 [0085]
- NAGAHARA et al. Nat. Med, 1998, vol. 4, 1449-1453 [0085]
- ROTHBARD et al. Nature Medicine, 2000, vol. 6 (11), 1253-1257 [0086]
- CROMBEZ et al. Biochem Soc. Trans., 2007, vol. 35, 44-46 [0086]
- TAKENOBU et al. Mol. Cancer Ther., 2002, vol. 1 (12), 1043-1049 [0086]
- SNYDER et al. PloS Biol., 2004, vol. 2, E36 [0086]
- MYOU et al. J. Immunol., 2003, vol. 171, 4399-4405 [0086]
- SHOKOLENKO et al. DNA Repair, 2005, vol. 4 (4), 511-518 [0087]
- SANTRA et al. Chem. Commun., 2005, 3144-3146 [0087]
- LIU et al. Biochem. And Biophys. Res. Comm., 2006,
- vol. 347 (1), 133-140 [0087]
- RAMSEY; FLYNN. Pharmacol Ther., 22 July 2015, vol. 15, S0163-7258 [0087]
- COKOL et al. EMBO Rep., 15 November 2000, vol.
- 1 (5), 411-415 [0088]
- FREITAS; CUNHA. Curr Genomics, December 2009, vol. 10 (8), 550-557 [0088]
- PALOMARES et al. Production of Recombinant Proteins: Challenges and Solutions. Methods Mol Biol., 2004, vol. 267, 15-52 [0090]
- LAFOUNTAINNE et al. Int J Pharm., 13 August 2015, vol. 494 (1), 180-194 [0090]
- SAMBROOK et al. Molecular Cloning, A Laboratory Manual, 2001 [0092]
- KRIEGLER. Gene Transfer and Expression: A Laboratory Manual, 1990 [0092]
- Current Protocols in Molecular Biology, 2010 [0092]
- PALVA et al. Gene, 1983, vol. 22, 229-235 [0092]
- GOSSEN; BUJARD. Proc. Natl. Acad. Sci. USA, 1992, vol. 89, 5547 [0093]
- OLIGINO et al. Gene Ther., 1998, vol. 5, 491-496 [0093]
- WANG et al. Gene Ther., 1997, vol. 4, 432-441 [0093]
- NEERING et al. Blood, 1996, vol. 88, 1147-55 [0093]
- RENDAHL et al. Nat. Biotechnol., 1998, vol. 16, 757-761 [0093]
- COLLEY et al. J. Biol. Chem., 1989, vol. 264, 17619-22 [0100]
- Guide to Protein Purification. Methods in Enzymology, 1990, vol. 182 [0100]
- MORRISON. J. Bacteriol., 1977, vol. 132, 349-351 [0100]

- CLARK-CURTISS; CURTISS et al. Methods in Enzymology. 1983, vol. 101, 347-362 [0100]
- ROHLAND et al. Genome Res., 2012, vol. 22, 939-46 [0108]
- REYON et al. Nat Biotechnol., 2012, vol. 30, 460-5 [0109]
- KLEINSTIVER et al. Nature, 2016, vol. 529, 490-5 [0110]
- PINELLO et al. Nat Biotechnol., 2016, vol. 34, 695-7 [0110]
- TSAI et al. Nat Biotechnol., 2015, vol. 33, 187-197 [0111] [0179]
- KLEINSTIVER et al. Nature, 2015, vol. 523, 481-5 [0111]
- TSAI et al. Nat Biotechnol., 2016, vol. 34, 483 [0111]
- KIM et al. Nat Biotechnol., 2016, vol. 34, 863-8 [0156]
- [0179]
- ZETSCHE et al. Cell, 2015, vol. 163, 759-71 [0156]
- FONFARA et al. Nature, 2016, vol. 532, 517-21 [0168]
- ZETSCHE et al. Nat Biotechnol., 2017, vol. 35, 31-34 [0168]
- TAK et al. Nat Methods, 2017, vol. 14, 1163-1166 [0168] [0183]
- ZETSCHE. Cell, 2015 [0171]
- KIM. Nature Biotechnology, 2016 [0171]
- KOMOR et al. Nature., 19 May 2016, vol. 533 (7603), 420-4 [0186]
- NISHIDA et al. Science, 16 September 2016, vol. 353 (6305) [0186]
- KIM et al. Nat Biotechnol., April 2017, vol. 35 (4), 371-376 [0186]
- CHEN et al. Science, 2018 [0188]
- EAST-SELETSKY. Nature, 2016, vol. 538, 270-273 [0188]