

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 029 182**

51 Int. Cl.:

C12N 9/22 (2006.01)

C12N 15/113 (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.04.2021** **PCT/US2021/028843**

87 Fecha y número de publicación internacional: **28.10.2021** **WO21217002**

96 Fecha de presentación y número de la solicitud europea: **23.04.2021** **E 21725880 (5)**

97 Fecha y número de publicación de la concesión europea: **19.02.2025** **EP 4118198**

54 Título: **Nucleasas guiadas por ARN y fragmentos activos y variantes de las mismas y métodos de uso**

30 Prioridad:

24.04.2020 US 202063014970 P

11.09.2020 US 202063077211 P

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

23.06.2025

73 Titular/es:

LIFE EDIT THERAPEUTICS, INC. (100.00%)

300 Morris Street, Suite 300

Durham, NC 27701, US

72 Inventor/es:

BOWEN, TYSON D.;

COYLE, MICHAEL;

CRAWLEY, ALEXANDRA BRINER y

ELICH, TEDD D.

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

Observaciones:

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

ES 3 029 182 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

Nucleasas guiadas por ARN y fragmentos activos y variantes de las mismas y métodos de uso

5 **DECLARACIÓN RELATIVA AL LISTADO DE SECUENCIAS**

El listado de secuencias asociado con la presente solicitud se proporciona en formato ASCII en lugar de una copia impresa, y por tanto se incorpora en la presente por referencia en la presente memoria descriptiva. La copia ASCII denominada L103438_1200WO_0084_1_SL tiene un tamaño de 878.362 bytes y se creó el 21 de abril de 2021.

10 **Campo de la invención**

La presente invención se refiere al campo de la biología molecular y la edición de genes, y se define por las reivindicaciones adjuntas.

15 **Antecedentes de la invención**

La edición o modificación dirigida del genoma se está convirtiendo rápidamente en una herramienta importante para la investigación básica y aplicada. Los métodos iniciales implicaban nucleasas diseñadas tales como meganucleasas, proteínas de fusión con dedos de zinc o TALEN, que requieren la generación de nucleasas quiméricas con dominios de unión al ADN específicos de secuencia programables y diseñados específicamente para cada secuencia diana concreta. Las nucleasas guiadas por ARN, tales como las proteínas asociadas a grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) (Cas) del sistema bacteriano CRISPR-Cas, permiten el direccionamiento a secuencias específicas mediante la formación de complejos de las nucleasas con ARN de guía que hibrida específicamente con una secuencia diana particular. Producir ARN de guía específicos de diana es menos costoso y más eficaz que generar nucleasas quiméricas para cada secuencia diana. Estas nucleasas guiadas por ARN pueden utilizarse para editar genomas, opcionalmente mediante la introducción de una rotura de la doble cadena específica de secuencia que se repara mediante la unión de extremos no homólogos (NHEJ), propensa a errores, para introducir una mutación en una localización genómica específica. Como alternativa, se puede introducir ADN heterólogo en el sitio genómico mediante reparación dirigida por homología. Las nucleasas guiadas por ARN (NGA) también pueden utilizarse para la edición de bases cuando se fusionan con una desaminasa.

Los documentos US 2019/264232 A1 y CN110551762 A divulgan cada uno una secuencia descrita como una enzima Cas9 que tiene una identidad de secuencia del 79,5 % con la presente SEQ ID NO: 83 en 1053 posiciones de aminoácidos.

35 **Breve resumen de la invención**

En un primer aspecto, la invención proporciona una molécula de ácido nucleico que comprende un polinucleótido que codifica un polipéptido de nucleasa guiada por ARN (NGA), en donde dicho polinucleótido comprende una secuencia de nucleótidos que codifica un polipéptido de NGA que comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83;

en donde dicho polipéptido de NGA se puede unir a una secuencia de ADN diana de una manera específica de secuencia guiada por ARN cuando se une a un ARN de guía (ARNg) capaz de hibridar con dicha secuencia de ADN diana y

en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.

En una realización, la molécula de ácido nucleico tiene codones optimizados para su expresión en una célula eucariota. En una realización, dicho polinucleótido que codifica un polipéptido NGA está unido operativamente a un promotor. En una realización, dicho promotor es heterólogo a dicho polinucleótido. En una realización, el polipéptido NGA está fusionado operativamente a un dominio efector. En una realización, el dominio efector comprende un dominio de corte, un dominio de desaminasa o un dominio modulador de expresión. En una realización, el dominio efector está fusionado operativamente en el extremo N-terminal, el extremo C-terminal o en una localización interna de dicho polipéptido NGA.

En un segundo aspecto, la invención proporciona un vector que comprende la molécula de ácido nucleico del primer aspecto. En una realización, el vector comprende además al menos una secuencia de nucleótidos que codifica dicho ARNg capaz de hibridarse con dicha secuencia de ADN diana.

En un tercer aspecto, la invención proporciona una célula que comprende la molécula de ácido nucleico del primer aspecto.

En un cuarto aspecto, la invención proporciona un polipéptido aislado de nucleasa guiada por ARN (NGA), en donde dicho polipéptido NGA comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de

secuencia con la SEQ ID NO: 83; y

en donde dicho polipéptido NGA se puede unir a una secuencia de ADN diana de una molécula de ADN de una manera específica de secuencia guiada por ARN cuando se une a un ARN de guía (ARNg) capaz de hibridar con dicha secuencia de ADN diana; y
 en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.

En una realización, el polipéptido NGA está fusionado operativamente a un dominio efector. En una realización, el dominio efector comprende un dominio de corte, un dominio de desaminasa o un dominio modulador de expresión. En una realización, el dominio efector está fusionado operativamente en el extremo N-terminal, el extremo C-terminal o en una localización interna de dicho polipéptido NGA.

En un quinto aspecto, la invención proporciona un sistema para unir una secuencia de ADN diana de una molécula de ADN, comprendiendo dicho sistema:

- a) uno o más ARN de guía capaces de hibridar con dicha secuencia de ADN diana o uno o más polinucleótidos que comprenden una o más secuencias de nucleótidos que codifican el uno o más ARN de guía (ARNg); y
- b) un polipéptido de nucleasa guiada por ARN (NGA) que comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83 o un polinucleótido que comprende una secuencia de nucleótidos que codifica el polipéptido de NGA;

en donde el uno o más ARN de guía pueden hibridar con la secuencia de ADN diana;
 en donde el uno o más ARN guía son capaces de formar un complejo con el polipéptido NGA para dirigir dicho polipéptido NGA a unirse a dicha secuencia de ADN diana de la molécula de ADN; y
 en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.

En una realización, al menos una de dichas secuencias nucleotídicas que codifican el uno o más ARN guía y dicha secuencia nucleotídica que codifica el polipéptido NGA está operativamente unida a un promotor heterólogo a dicha secuencia nucleotídica. En una realización, dicha secuencia de ADN diana es una secuencia de ADN diana eucariota. En una realización, las secuencias de nucleótidos que codifican el uno o más ARN guía y la secuencia de nucleótidos que codifica un polipéptido NGA se encuentran en un vector. En una realización, dicho ARNg se selecciona entre:

- a) un ARNg que comprenda una secuencia repetida CRISPR que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 84 y un ARNtrat que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 85; y
- b) un ARNg que comprenda una secuencia repetida CRISPR que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 84 y un ARNtrat que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 78.

En una realización del primer, cuarto o quinto aspectos, dicho polipéptido NGA es una nucleasa inactiva o es una nickasa. En una realización del primer, cuarto o quinto aspectos, el polipéptido de NGA está fusionado operativamente a un polipéptido de edición de bases. En una realización, el polipéptido de edición de bases es una desaminasa. En una realización del primer, cuarto o quinto aspectos, el polipéptido NGA comprende una o más señales de localización nuclear.

En un sexto aspecto, la invención proporciona un método *in vitro* o *ex vivo* para unir una secuencia de ADN diana de una molécula de ADN que comprende suministrar un sistema según el quinto aspecto, a dicha secuencia de ADN diana o a una célula que comprende la secuencia de ADN diana, en donde el método no es un proceso para modificar la identidad genética de la línea germinal de un ser humano.

En un séptimo aspecto, la invención proporciona un método *in vitro* o *ex vivo* para escindir y/o modificar una secuencia de ADN diana de una molécula de ADN, que comprende poner en contacto la molécula de ADN con:

- a) un polipéptido de nucleasa guiada por ARN (NGA), en donde dicha NGA comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83; y
- b) uno o más ARN de guía capaces de dirigir la NGA de (a) a la secuencia de ADN diana;

en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82;
 en donde el uno o más ARN de guía hibridan con la secuencia de ADN diana, dirigiendo así dicho polipéptido de NGA para que se una a dicha secuencia de ADN diana y se produzca la escisión y/o modificación de dicha secuencia de ADN diana, y
 en donde el método no es un proceso para modificar la identidad genética de la línea germinal de un ser humano.

En una realización, la escisión por dicho polipéptido NGA genera una rotura bicatenaria o una rotura monocatenaria. En una realización, dicha secuencia de ADN diana modificada comprende:

- a) delección de al menos un nucleótido de la secuencia de ADN diana; o
- b) mutación de al menos un nucleótido en la secuencia de ADN diana.

En una realización del primer, segundo, tercer, cuarto, quinto, sexto o séptimo aspectos, dicho polipéptido de NGA comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 83.

- Se proporcionan composiciones y métodos para unir una secuencia diana de interés. Las composiciones encuentran uso en la escisión o modificación de una secuencia diana de interés, la detección de una secuencia diana de interés y la modificación de la expresión de una secuencia de interés. Las composiciones comprenden polipéptidos de nucleasas guiadas por ARN (NGA), los ARN de CRISPR (los ARNcr), los ARN transactivadores de CRISPR (los ARNtrat), los ARN de guía (ARNg), moléculas de ácido nucleico que codifican los mismos, y vectores y células hospedadoras que comprenden las moléculas de ácido nucleico. También se proporcionan sistemas de NGA para la unión a una secuencia diana de interés, en donde el sistema de NGA comprende un polipéptido de nucleasa guiada por ARN y uno o más ARN de guía. Por tanto, los métodos divulgados en el presente documento están destinados a la unión con una secuencia diana de interés y, en algunas realizaciones, escindir o modificar la secuencia diana de interés. La secuencia diana de interés se puede modificar, por ejemplo, como resultado de la unión de extremos no homólogos, reparación dirigida por homología con una secuencia donante introducida, o edición de bases.

Las referencias a métodos de tratamiento mediante terapia con una composición, sistema u otro agente deben interpretarse como referencias a la composición, sistema u otro agente para su uso en dichos métodos.

Descripción detallada

I. Descripción general

- Las nucleasas guiadas por ARN (NGA) permiten la manipulación dirigida de uno o más sitios específicos dentro de un genoma y son útiles en el contexto del direccionamiento a genes para aplicaciones terapéuticas y de investigación. En una diversidad de organismos, incluyendo mamíferos, las nucleasas guiadas por ARN se han utilizado para el diseño técnico del genoma mediante la estimulación de la unión de extremos no homólogos y la recombinación homóloga, por ejemplo. Las composiciones y métodos descritos en el presente documento son útiles para crear roturas monocatenarias o bicatenarias en polinucleótidos, para modificar polinucleótidos, detectar un sitio particular dentro de un polinucleótido o modificar la expresión de un gen particular.

- Las nucleasas guiadas por ARN divulgadas en el presente documento pueden alterar la expresión génica modificando una secuencia diana. En realizaciones específicas, las nucleasas guiadas por ARN se dirigen a la secuencia diana mediante un ARN de guía (ARNg) como parte de un sistema de nucleasa guiada por ARN de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR). Las NGA se consideran "guiadas por ARN" porque los ARN guía forman un complejo con las nucleasas guiadas por ARN para dirigir a la nucleasa guiada por ARN a unirse a una secuencia diana y en algunas realizaciones, introducir una rotura monocatenaria o bicatenaria en la secuencia diana. Una vez escindida la secuencia diana, la rotura se puede reparar de manera que la secuencia de ADN de la secuencia diana se modifique durante el proceso de reparación. Por tanto, en el presente documento se proporcionan métodos para usar las nucleasas guiadas por ARN para modificar una secuencia diana en el ADN de las células hospedadoras. Por ejemplo, se pueden usar nucleasas guiadas por ARN para modificar una secuencia diana en un locus genómico de células eucariotas o células procariotas.

II. Nucleasas guiadas por ARN

- En el presente documento se proporcionan nucleasas guiadas por ARN según el cuarto aspecto. El término nucleasa guiada por ARN (NGA) se refiere a un polipéptido que se une a una secuencia de nucleótidos diana particular de una manera específica de secuencia y se dirige a la secuencia de nucleótidos diana mediante una molécula de ARN de guía que forma un complejo con el polipéptido e hibrida con la secuencia diana. Aunque una nucleasa guiada por ARN puede ser capaz de escindir la secuencia diana al unirse, la expresión nucleasa guiada por ARN también abarca nucleasas guiadas por ARN sin actividad de nucleasa que son capaces de unirse a, pero no de escindir, una secuencia diana. La escisión de una secuencia diana mediante una nucleasa guiada por ARN puede dar como resultado una rotura monocatenaria o bicatenaria. Las nucleasas guiadas por ARN que sólo son capaces de escindir una única cadena de una molécula de ácido nucleico bicatenario se denominan en el presente documento nickasas.

- Las nucleasas guiadas por ARN divulgadas en el presente documento incluyen la nucleasa guiada por ARN APG01604, cuya secuencia de aminoácidos se expone como la SEQ ID NO: 83, y fragmentos o variantes activas de la misma que conserven la capacidad de unirse a una secuencia de nucleótidos diana de una manera específica de secuencia guiada por ARN. En algunas de estas realizaciones, el fragmento o variante activa de la NGA APG01604 es capaz de escindir una secuencia diana monocatenaria o bicatenaria. La variante activa de la NGA APG01604 comprende una secuencia de aminoácidos que tiene al menos el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el

96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia con la secuencia de aminoácidos expuesta en la SEQ ID NO: 83. En determinadas realizaciones, un fragmento activo de la NGA APG01604 comprende al menos 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 o más restos de aminoácidos contiguos de la secuencia de aminoácidos expuesta como la SEQ ID NO: 83. Las nucleasas guiadas por
 5 ARN proporcionadas en el presente documento pueden comprender al menos un dominio de nucleasa (por ejemplo, dominio de ADNasa, ARNasa) y al menos el dominio de reconocimiento de ARN y/o de unión a ARN para interactuar con los ARN de guía. Otros dominios que se pueden encontrar en las nucleasas guiadas por ARN proporcionadas en el presente documento incluyen, pero sin limitación: dominios de unión a ADN, dominios de helicasa, dominios de interacción proteína-proteína y dominios de dimerización. En realizaciones específicas, las nucleasas guiadas por ARN
 10 proporcionadas en el presente documento pueden comprender al menos el 70 %, el 75 %, el 80 %, el 85 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia con uno o más de un dominio de unión al ADN, dominio de helicasa, dominio de interacción proteína-proteína y dominio de dimerización.

15 Una secuencia de nucleótidos diana se une a una nucleasa guiada por ARN proporcionada en el presente documento e hibrida con el ARN de guía asociado con la nucleasa guiada por ARN. La secuencia diana puede escindirse posteriormente mediante la nucleasa guiada por ARN si el polipéptido posee actividad de nucleasa. Los términos "escindir" o "escisión" se refieren a la hidrólisis de al menos el enlace fosfodiéster dentro de la cadena principal de una secuencia de nucleótidos diana, lo que puede dar como resultado roturas monocatenarias o bicatenarias dentro de la
 20 secuencia diana. Las NGA divulgadas en el presente documento pueden escindir nucleótidos dentro de un polinucleótido, funcionando como una endonucleasa, o pueden ser una exonucleasa, que elimina nucleótidos sucesivos del extremo (el extremo 5' y/o el extremo 3') de un polinucleótido. En otras realizaciones, las NGA divulgadas pueden escindir nucleótidos de una secuencia diana dentro de cualquier posición de un polinucleótido y, por tanto, funcionar como endonucleasa y exonucleasa. La escisión de un polinucleótido diana mediante las NGA divulgadas en
 25 el presente documento puede dar como resultado roturas escalonadas o extremos romos.

Las nucleasas guiadas por ARN divulgadas en el presente documento pueden ser secuencias de tipo silvestre procedentes de especies bacterianas o de arqueas. Como alternativa, las nucleasas guiadas por ARN pueden ser
 30 variantes o fragmentos de polipéptidos de tipo silvestre. La NGA de tipo silvestre se puede modificar para alterar la actividad de la nucleasa o alterar la especificidad de PAM, por ejemplo. En algunas realizaciones, la nucleasa guiada por ARN no es de origen natural.

En determinadas realizaciones, la nucleasa guiada por ARN funciona como una nickasa, escindiendo sólo una única cadena de la secuencia de nucleótidos diana. Dichas nucleasas guiadas por ARN tienen un único dominio de nucleasa funcional. En realizaciones particulares, la nickasa es capaz de escindir la hebra positiva o la hebra negativa. En algunas de estas realizaciones, se han mutado dominios de nucleasa adicionales de manera que la actividad de
 35 nucleasa se reduce o elimina.

En otras realizaciones, la nucleasa guiada por ARN carece totalmente de actividad nucleasa y se denomina en el
 40 presente documento nucleasa muerta o nucleasa inactiva. Cualquier método conocido en la técnica para introducir mutaciones en una secuencia de aminoácidos, tal como la mutagénesis mediada por PCR y la mutagénesis dirigida al sitio, puede utilizarse para generar nickasas o NGA muertos por nucleasas. Véase, por ejemplo, las publicaciones de EE. UU. N.º 2014/0068797 y la patente de EE.UU. N.º 9.790.490.

45 Las nucleasas guiadas por ARN que carecen de actividad nucleasa pueden utilizarse para liberar un polipéptido fusionado, polinucleótido o carga útil de molécula pequeña a una localización genómica concreta. En algunas de estas realizaciones, el polipéptido de NGA o ARN de guía se puede fusionar con un marcador de detección para permitir la detección de una secuencia particular. Como ejemplo no limitante, una NGA con actividad de nucleasa se puede
 50 fusionar a un marcador de detección (por ejemplo, una proteína fluorescente) y dirigirse a una secuencia particular asociada con una enfermedad para permitir la detección de la secuencia asociada a la enfermedad.

Como alternativa, las NGA sin actividad de nucleasa pueden dirigirse a ubicaciones genómicas particulares para
 55 alterar la expresión de una secuencia deseada. En algunas realizaciones, la unión de una nucleasa guiada por ARN sin actividad de nucleasa a una secuencia diana da como resultado la reducción de la expresión de la secuencia diana o de un gen bajo control transcripcional por la secuencia diana al interferir con la unión de la ARN polimerasa o de factores de transcripción dentro de la región genómica diana. En otras realizaciones, la NGA (por ejemplo, una NGA sin actividad de nucleasa) o su ARN de guía formando complejo comprende además un modulador de la expresión que, al unirse a una secuencia diana, sirve para reprimir o activar la expresión de la secuencia diana o un gen bajo control transcripcional por la secuencia diana. En algunas de estas realizaciones, el modulador de expresión modula
 60 la expresión de la secuencia diana o del gen regulado a través de mecanismos epigenéticos.

En otras realizaciones, las NGA sin actividad de nucleasa o una NGA con solo actividad de nickasa pueden dirigirse a ubicaciones genómicas particulares para modificar la secuencia de un polinucleótido diana mediante fusión con un
 65 polipéptido de edición de bases, por ejemplo, un polipéptido de desaminasa o una variante activa o un fragmento del mismo, que modifica químicamente de forma directa (por ejemplo, desamina) una nucleobase, dando como resultado la conversión de una base nucleotídica en otra. El polipéptido de edición de bases se puede fusionar al NGA en su

extremo aminoterminal o carboxiterminal. Además, el polipéptido de edición de bases puede fusionarse a la NGA mediante un enlazador peptídico. Un ejemplo no limitante de un polipéptido de desaminasa que es útil para tales composiciones y métodos incluye el editor de bases citidina desaminasa o una adenosina desaminasa (tal como el editor de bases de adenina desaminasa descrito en Gaudelli. *et al.* (2017) *Nature* 551:464-471, las publicaciones de EE. UU. N.º 2017/0121693 y 2018/0073012, y la publ. internacional. N.º WO 2018/027078, o cualquiera de las desaminasas divulgadas en la publicación internacional N.º WO 2020/139873). Además, es conocido en la técnica que ciertas proteínas de fusión entre un NGA y una enzima editora de bases también pueden comprender al menos un polipéptido estabilizador del uracilo que aumenta la tasa de mutación de una citidina, desoxicitidina o citosina a una timidina, desoxitimidina o timina en una molécula de ácido nucleico mediante una desaminasa. Ejemplos no limitantes de polipéptidos estabilizadores de uracilo incluyen los divulgados en la solicitud provisional de Estados Unidos n.º 63/052.175, presentada el 15 de julio de 2020, incluyendo USP2 (SEQ ID NO: 1089) y un dominio inhibidor de uracilo glicosilasa (UGI) (SEQ ID NO: 212), que puede aumentar la eficiencia de edición de bases. Por lo tanto, una proteína de fusión puede comprender un NGA descrito en el presente documento o una variante del mismo, una desaminasa y opcionalmente al menos un polipéptido estabilizador del uracilo, tal como UGI o USP2. En determinadas realizaciones, el NGA fusionado al polipéptido de edición de bases es una nickasa que corta la cadena de ADN sobre la que no actúa el polipéptido de edición de bases (por ejemplo, una desaminasa).

Las nucleasas guiadas por ARN que están fusionadas a un polipéptido o dominio pueden estar separadas o unidas mediante un enlazador. El término "enlazador", como se utiliza en el presente documento, se refiere a un grupo químico o una molécula que une dos moléculas o restos, por ejemplo, un dominio de unión y un dominio de escisión de una nucleasa. En algunas realizaciones, un enlazador une un dominio de unión a ARNg de una nucleasa guiada por ARN y un polipéptido de edición de bases, tal como una desaminasa. En algunas realizaciones, un enlazador une una NGA sin actividad de nucleasa y una desaminasa. Típicamente, el enlazador está colocado entre, o flanqueado por, dos grupos, moléculas u otros restos, y se conecta a cada uno de ellos mediante un enlace covalente, conectándolos así. En algunas realizaciones, el enlazador es un aminoácido o una pluralidad de aminoácidos (por ejemplo, un péptido o proteína). En algunas realizaciones, el enlazador es una molécula orgánica, un grupo, polímero o resto químico. En algunas realizaciones, el enlazador tiene 5-100 aminoácidos de longitud, por ejemplo, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 o 150-200 aminoácidos de longitud. También se contemplan enlazadores más largos o más cortos.

Las nucleasas guiadas por ARN divulgadas en el presente documento pueden comprender al menos una señal de localización nuclear (SLN) para potenciar el transporte de la NGA al núcleo de una célula. Las señales de localización nuclear se conocen en la técnica y generalmente comprenden un tramo de aminoácidos básicos (véase, por ejemplo, Lange *et al.*, *J. Biol. Chem.* (2007) 282:5101-5105). En realizaciones particulares, la NGA comprende 2, 3, 4, 5, 6 o más señales de localización nuclear. La señal o señales de localización nuclear pueden ser una SLN heteróloga. Los ejemplos no limitantes de señales de localización nuclear útiles para las NGA divulgadas en el presente documento son las señales de localización nuclear del antígeno T grande de SV40, nucleoplasmina y c-Myc (véase, por ejemplo, Ray *et al.* (2015) *Bioconjug Chem* 26(6):1004-7). En realizaciones particulares, la NGA comprende la secuencia de SLN expuesta como SEQ ID NO: 251 o 253. La NGA puede comprender una o más secuencias de SLN en su extremo N, extremo C o tanto en el extremo N como en el extremo C. Por ejemplo, la NGA puede comprender dos secuencias de SLN en la región aminoterminal y cuatro secuencias de SLN en la región carboxiterminal.

También pueden utilizarse otras secuencias de señal de localización conocidas en la técnica que localizan polipéptidos en una localización o localizaciones subcelulares particulares para dirigir las NGA, incluyendo, aunque no de forma limitativa, secuencias de localización de plástidos, secuencias de localización mitocondrial y secuencias señal de doble direccionamiento que se dirigen tanto al plástido como a las mitocondrias (véase, por ejemplo, Nassoury y Morse (2005) *Biochim Biophys Acta* 1743:5-19; Kunze y Berger (2015) *Front Physiol* dx.doi.org/10.3389/fphys.2015.00259; Herrmann y Neupert (2003) *IUBMB Life* 55:219-225; Soll (2002) *Curr Opin Plant Biol* 5:529-535; Carrie y Small (2013) *Biochim Biophys Acta* 1833:253-259; Carrie *et al.* (2009) *FEBS J* 276:1187-1195; Silva-Filho (2003) *Curr Opin Plant Biol* 6:589-595; Peeters y Small (2001) *Biochim Biophys Acta* 1541:54-63; Murcha *et al.* (2014) *J Exp Bot* 65:6301-6335; Mackenzie (2005) *Trends Cell Biol* 15:548-554; Glaser *et al.* (1998) *Plant Mol Biol* 38:311-338).

En determinadas realizaciones, las nucleasas guiadas por ARN divulgadas en el presente documento comprenden al menos el dominio de penetración celular que facilita la captación celular de la NGA. Los dominios de penetración celular, son conocidos en la técnica y generalmente comprenden tramos de restos de aminoácido cargados positivamente (es decir, dominios policatiónicos de penetración celular), que alternan restos de aminoácidos polares y restos de aminoácidos no polares (es decir, dominios de penetración de células anfipáticos), o restos de aminoácidos hidrófobos (es decir, dominios hidrófobos de penetración celular) (véase, por ejemplo, Milletti F. (2012) *Drug Discov Today* 17:850-860). Un ejemplo no limitante de un dominio de penetración celular es el activador transcripcional transactivador (TAT) del virus de la inmunodeficiencia humana 1.

La señal de localización nuclear, la señal de localización de plástidos, la señal de localización mitocondrial, la señal de localización de doble direccionamiento y/o el dominio de penetración celular pueden ubicarse en el extremo amino (extremo N), el extremo carboxilo (extremo C) o en una ubicación interna de la nucleasa guiada por ARN.

Los NGA actualmente divulgados pueden estar fusionados con un dominio efector, tal como un dominio de escisión,

un dominio de desaminasa o un dominio modulador de expresión, ya sea directa o indirectamente a través de un péptido enlazador. Dicho dominio puede ubicarse en el extremo N, el extremo C o una ubicación interna de la nucleasa guiada por ARN. En algunas de estas realizaciones, el componente de NGA de la proteína de fusión es una NGA sin actividad de nucleasa.

En algunas realizaciones, la proteína de fusión de NGA comprende un dominio de escisión, que es cualquier dominio que sea capaz de escindir un polinucleótido (es decir, ARN, ADN o un híbrido de ARN/ADN) e incluye, pero sin limitación, endonucleasas de restricción y endonucleasas de asentamiento, tales como las endonucleasas de tipo IIS (por ejemplo, *FokI*) (véase, por ejemplo, Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Linn *et al.* (eds.)

Nucleases, Cold Spring Harbor Laboratory Press, 1993). En otras realizaciones, la proteína de fusión de NGA comprende un dominio de desaminasa que desamina una base nucleotídica, dando como resultado la conversión de una nucleobase en otra, e incluye, pero sin limitación, un editor de bases de citidina desaminasa o de adenosina desaminasa (véase, por ejemplo, Gaudelli *et al.* (2017) *Nature* 551:464-471, las publicaciones de EE. UU. N.º 2017/0121693 y 2018/0073012, patente de Estados Unidos n.º 9.840.699 y la publicación internacional N.º WO/2018/027078).

En algunas realizaciones, el dominio efector de la proteína de fusión de NGA puede ser un dominio modulador de la expresión, que es un dominio que sirve para regular al alza o a la baja la transcripción. El dominio modulador de la expresión puede ser un dominio de modificación epigenética, un dominio represor transcripcional o un dominio de activación transcripcional.

En algunas de estas realizaciones, el modulador de la expresión de la proteína de fusión de NGA comprende un dominio de modificación epigenética que modifica covalentemente el ADN o las proteínas histonas para alterar la estructura de las histonas y/o la estructura cromosómica sin alterar la secuencia del ADN, conduciendo a cambios en la expresión genética (es decir, regulación al alza o a la baja). Los ejemplos no limitantes de modificaciones epigenéticas incluyen acetilación o metilación de restos lisina, metilación de arginina, fosforilación de serina y treonina, y ubiquitinación y sumoilación de lisina de proteínas histonas, y metilación e hidroximetilación de restos citosina en el ADN. Los ejemplos no limitantes de dominios de modificación epigenética incluyen dominios de histona acetiltransferasa, dominios de histona desacetilasa, dominios de histona metiltransferasa, dominios de histona desmetilasa, dominios de ADN metiltransferasa y dominios de ADN desmetilasa.

En otras realizaciones, el modulador de la expresión de la proteína de fusión comprende un dominio represor transcripcional, que interactúa con elementos de control transcripcional y/o proteínas reguladoras transcripcionales, tales como ARN polimerasas y factores de transcripción, para reducir o terminar la transcripción de al menos el gen. Los dominios represores transcripcionales se conocen en la técnica e incluyen, pero sin limitación, represores de tipo Sp1, IκB y dominios de caja asociados a Krüppel (KRAB).

En otras realizaciones más, el modulador de la expresión de la proteína de fusión comprende un dominio de activación transcripcional, que interactúa con elementos de control transcripcional y/o proteínas reguladoras transcripcionales, tales como ARN polimerasas y factores de transcripción, para aumentar o activar la transcripción de al menos el gen. Los dominios de activación transcripcional se conocen en la técnica e incluyen, pero sin limitación, un dominio de activación de la VP16 del virus del herpes simple y un dominio de activación NFAT.

Los polipéptidos de NGA divulgados en el presente documento pueden comprender un marcador detectable o una etiqueta de purificación. El marcador de detección o la etiqueta de purificación puede ubicarse en el extremo N, el extremo C o una ubicación interna de la nucleasa guiada por ARN, ya sea directa o indirectamente a través de un péptido enlazador. En algunas de estas realizaciones, el componente de NGA de la proteína de fusión es una NGA sin actividad de nucleasa. En otras realizaciones, el componente de NGA de la proteína de fusión es una NGA con actividad de nickasa.

Un marcador de detección es una molécula que puede visualizarse u observarse de otro modo. El marcador detectable puede fusionarse a la NGA como una proteína de fusión (por ejemplo, proteína fluorescente) o puede ser una molécula pequeña conjugada al polipéptido de NGA que puede detectarse visualmente o por otros medios. Los marcadores detectables que pueden fusionarse con las NGA divulgadas en el presente documento como una proteína de fusión incluyen cualquier dominio proteico detectable, incluyendo, pero sin limitación, una proteína fluorescente o un dominio proteico que puede detectarse con un anticuerpo específico. Los ejemplos no limitantes de proteínas fluorescentes incluyen proteínas verdes fluorescentes (por ejemplo, GFP, EGFP, ZsGreen1) y proteínas fluorescentes amarillas fluorescentes (por ejemplo, YFP, EYFP, ZsYellow1). Los ejemplos no limitantes de marcadores detectables de molécula pequeña incluyen marcadores radiactivos, tales como ³H y ³⁵S.

Los polipéptidos NGA también pueden incluir una etiqueta de purificación, que es cualquier molécula que se puede utilizar para aislar una proteína o proteína fusionada de una mezcla (por ejemplo, una muestra biológica, un medio de cultivo). Los ejemplos no limitantes de etiquetas de purificación incluyen biotina, myc, proteína de unión a maltosa (MBP), glutatión-S-transferasa (GST) y etiqueta 3X FLAG.

II. ARN de guía

La presente divulgación proporciona ARN de guía y polinucleótidos que los codifican. La expresión "ARN de guía" se refiere a una secuencia de nucleótidos que tiene suficiente complementariedad con una secuencia de nucleótidos diana para hibridar con la secuencia diana y dirigir la unión específica de secuencia de una nucleasa guiada por ARN asociada a la secuencia de nucleótidos diana. Por tanto, el ARN de guía respectivo de una NGA es una o más moléculas de ARN (generalmente, una o dos), que pueden unirse a la NGA y guiar a la NGA para que se una a una secuencia de nucleótidos diana particular, y en las realizaciones en donde la NGA tiene actividad de nickasa o nucleasa, también escinde la secuencia de nucleótidos diana. En general, un ARN de guía comprende un ARN de CRISPR (ARNcr) y un ARN transactivador de CRISPR (ARNtrac). Los ARN de guía nativos que comprenden tanto un ARNcr como un ARNtrac generalmente comprenden dos moléculas de ARN separadas que hibridan entre sí a través de la secuencia de repetición del ARNcr y la secuencia antirrepetición del ARNtrac.

Las secuencias nativas de repetición directa dentro de una agrupación de CRISPR generalmente varían en longitud de 28 a 37 pares de bases, aunque la longitud puede variar entre aproximadamente 23 pb y aproximadamente 55 pb. Las secuencias espaciadoras dentro de una agrupación de CRISPR generalmente varían de aproximadamente 32 a aproximadamente 38 pb de longitud, aunque la longitud puede ser de entre aproximadamente 21 pb y aproximadamente 72 pb. Cada agrupación de CRISPR generalmente comprende menos de 50 unidades de la secuencia de repetición-espaciador de CRISPR. Los CRISPR se transcriben como parte de un transcrito largo denominado transcrito de CRISPR primario, que comprende gran parte de la agrupación de CRISPR. El transcrito de CRISPR primario es escindido por proteínas Cas para producir ARNcr o, en algunos casos, para producir pre-ARNcr que son procesados adicionalmente por proteínas Cas adicionales en ARNcr maduros. Los ARNcr maduros comprenden una secuencia espaciadora y una secuencia de repetición de CRISPR. En algunas realizaciones en las que los pre-ARNcr se procesan en ARNcr maduros (o procesados), la maduración implica la eliminación de aproximadamente uno a aproximadamente seis o más nucleótidos 5', 3' o 5' y 3'. A los efectos de editar el genoma o dirigirse a una secuencia de nucleótidos diana particular de interés, estos nucleótidos que se eliminan durante la maduración de la molécula de pre-ARNcr no son necesarios para generar o diseñar un ARN de guía.

Un ARN de CRISPR (ARNcr) comprende una secuencia espaciadora y una secuencia de repetición de CRISPR. La "secuencia espaciadora" es la secuencia de nucleótidos que hibrida directamente con la secuencia de nucleótidos diana de interés. La secuencia espaciadora está técnicamente diseñada para que sea total o parcialmente complementaria con la secuencia diana de interés. En diferentes realizaciones, la secuencia espaciadora puede comprender de aproximadamente 8 nucleótidos a aproximadamente 30 nucleótidos, o más. Por ejemplo, la secuencia espaciadora puede ser de aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12, aproximadamente 13, aproximadamente 14, aproximadamente 15, aproximadamente 16, aproximadamente 17, aproximadamente 18, aproximadamente 19, aproximadamente 20, aproximadamente 21, aproximadamente 22, aproximadamente 23, aproximadamente 24, aproximadamente 25, aproximadamente 26, aproximadamente 27, aproximadamente 28, aproximadamente 29, aproximadamente 30 o más nucleótidos de longitud. En algunas realizaciones, la secuencia espaciadora tiene 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, o más nucleótidos de longitud. En algunas realizaciones, la secuencia espaciadora es de aproximadamente 10 a aproximadamente 26 nucleótidos de longitud, o de aproximadamente 12 a aproximadamente 30 nucleótidos de longitud. En realizaciones particulares, la secuencia espaciadora tiene aproximadamente 30 nucleótidos de longitud. En algunas realizaciones, la secuencia espaciadora tiene 30 nucleótidos de longitud. En algunas realizaciones, el grado de complementariedad entre una secuencia espaciadora y su secuencia diana correspondiente, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, se sitúa entre el 50 % y el 99 % o más, incluyendo, pero sin limitaciones, aproximadamente o más de aproximadamente 50 %, aproximadamente 60 %, aproximadamente 70 %, aproximadamente 75 %, aproximadamente 80 %, aproximadamente 81 %, aproximadamente 82 %, aproximadamente 83 %, aproximadamente 84 %, aproximadamente 85 %, aproximadamente 86 %, aproximadamente 87 %, aproximadamente 88 %, aproximadamente 89 %, aproximadamente 90 %, aproximadamente 91 %, aproximadamente 92 %, aproximadamente 93 %, aproximadamente 94 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, aproximadamente el 99 % o más. En realizaciones particulares, el grado de complementariedad entre una secuencia espaciadora y su secuencia diana correspondiente, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, es del 50 %, el 60 %, el 70 %, el 75 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más. En realizaciones particulares, la secuencia espaciadora está libre de una estructura secundaria, lo que puede predecirse usando cualquier algoritmo de plegamiento de polinucleótidos adecuado conocido en la técnica, incluyendo, pero sin limitación, mFold (véase, por ejemplo, Zuker y Stiegler (1981) *Nucleic Acids Res.* 9:133-148) y RNAfold (véase, por ejemplo, Gruber *et al.* (2008) *Cell* 106(1):23-24).

La secuencia de repetición de ARN CRISPR comprende una secuencia de nucleótidos que forma una estructura, ya sea por sí sola o junto con un ARNtrac hibridado, que es reconocido por la molécula de NGA. En diferentes realizaciones, la secuencia de repetición del ARN de CRISPR puede comprender de aproximadamente 8 nucleótidos a aproximadamente 30 nucleótidos, o más. Por ejemplo, la secuencia de repetición de CRISPR puede ser de aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12, aproximadamente 13, aproximadamente 14, aproximadamente 15, aproximadamente 16, aproximadamente 17,

aproximadamente 18, aproximadamente 19, aproximadamente 20, aproximadamente 21, aproximadamente 22, aproximadamente 23, aproximadamente 24, aproximadamente 25, aproximadamente 26, aproximadamente 27, aproximadamente 28, aproximadamente 29, aproximadamente 30 o más nucleótidos de longitud. En realizaciones particulares, la secuencia de repetición CRISPR tiene 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más nucleótidos de longitud. En algunas realizaciones, el grado de complementariedad entre una secuencia de repetición de CRISPR y su secuencia de ARNtrac correspondiente, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, es de aproximadamente o más de aproximadamente el 50 %, aproximadamente 60 %, aproximadamente 70 %, aproximadamente 75 %, aproximadamente 80 %, aproximadamente 81 %, aproximadamente 82 %, aproximadamente 83 %, aproximadamente 84 %, aproximadamente 85 %, aproximadamente 86 %, aproximadamente 87 %, aproximadamente 88 %, aproximadamente 89 %, aproximadamente 90 %, aproximadamente 91 %, aproximadamente 92 %, aproximadamente 93 %, aproximadamente 94 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, aproximadamente el 99 % o más. En realizaciones particulares, el grado de complementariedad entre una secuencia de repetición de CRISPR y su secuencia de ARNtrac correspondiente, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, es del 50 %, el 60 %, el 70 %, el 75 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más.

En realizaciones particulares, la secuencia de repetición de CRISPR comprende la secuencia de nucleótidos de la SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124, o una variante activa o fragmento del mismo que cuando está comprendido dentro de un ARN guía, es capaz de dirigir la unión específica de secuencia de una nucleasa guiada por ARN asociada proporcionada en el presente documento a una secuencia diana de interés. En determinadas realizaciones, una variante de secuencia de repetición de CRISPR activa de una secuencia de tipo silvestre comprende una secuencia de nucleótidos que tiene al menos el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 85 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia con la secuencia de nucleótidos expuesta como las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124. En determinadas realizaciones, un fragmento de secuencia de repetición de CRISPR activo de una secuencia de tipo silvestre comprende al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 o 22 nucleótidos contiguos de la secuencia de nucleótidos expuesta como las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124.

En determinadas realizaciones, el ARNcr no es de origen natural. En algunas de estas realizaciones, la secuencia de repetición de CRISPR específica no está unida a la secuencia espaciadora técnicamente diseñada en la naturaleza y la secuencia de repetición de CRISPR se considera heteróloga para la secuencia espaciadora. En determinadas realizaciones, la secuencia espaciadora es una secuencia diseñada técnicamente que no es de origen natural.

Una molécula de ARN transactivador de CRISPR o ARNtrac comprende una secuencia de nucleótidos que comprende una región que tiene suficiente complementariedad para hibridar con una secuencia de repetición de CRISPR de un ARNcr, que se denomina en el presente documento región antirrepetición. En algunas realizaciones, la molécula de ARNtrac comprende además una región con una estructura secundaria (por ejemplo, tallo-bucle) o forma una estructura secundaria al hibridar con su ARNcr correspondiente. En realizaciones particulares, la región del ARNtrac que es total o parcialmente complementaria a una secuencia de repetición de CRISPR está en el extremo 5' de la molécula y el extremo 3' del ARNtrac comprende una estructura secundaria. Esta región de estructura secundaria comprende generalmente varias estructuras en horquilla, incluyendo la horquilla nexa, que se encuentra adyacente a la secuencia antirrepetición. El nexa forma el núcleo de las interacciones entre el ARN guía y la NGA y se encuentra en la intersección entre el ARN guía, la NGA y el ADN diana. La horquilla nexa a menudo tiene una secuencia de nucleótidos conservada en la base del tallo de la horquilla, con el motivo UNANNC (SEQ ID NO: 132) que se encuentra en muchas horquillas de nexa en los ARNtrac. Cabe destacar que, varias de las NGA de la invención utilizan ARNtrac que comprenden secuencias no canónicas en la base del tallo de la horquilla de sus horquillas nexa, incluyendo UNANNA, UNANNU, UNANNG y CNANNC (SEQ ID NO: 129, 130, 131 y 133, respectivamente). A menudo hay horquillas terminales en el extremo 3' del ARNtrac que pueden variar en estructura y número, pero a menudo comprenden una horquilla terminadora transcripcional independiente de Rho rica en GC seguida de una sucesión de U en el extremo 3'. Véase, por ejemplo, Briner *et al.* (2014) Molecular Cell 56:333-339, Briner y Barrangou (2016) Cold Spring Harb Protoc; doi: 10.1101/pdb.top090902 y la publicación de Estados Unidos n.º 2017/0275648.

En diferentes realizaciones, la región antirrepetición del ARNtrac que es total o parcialmente complementaria a la secuencia de repetición de CRISPR comprende de aproximadamente 8 nucleótidos a aproximadamente 30 nucleótidos, o más. Por ejemplo, la región de emparejamiento de bases entre la secuencia antirrepetición de ARNtrac y la secuencia de repetición de CRISPR puede ser aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12, aproximadamente 13, aproximadamente 14, aproximadamente 15, aproximadamente 16, aproximadamente 17, aproximadamente 18, aproximadamente 19, aproximadamente 20, aproximadamente 21, aproximadamente 22, aproximadamente 23, aproximadamente 24, aproximadamente 25, aproximadamente 26, aproximadamente 27, aproximadamente 28, aproximadamente 29, aproximadamente 30 o más nucleótidos de longitud. En realizaciones particulares, la región de emparejamiento de bases entre la secuencia de anti-repetición de ARNtrac y la secuencia de repetición de CRISPR tiene 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más nucleótidos de longitud. En algunas realizaciones, el grado de

complementariedad entre una secuencia de repetición de CRISPR y su correspondiente secuencia antirrepetición de ARNtrat, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, es de aproximadamente o más de aproximadamente el 50 %, aproximadamente 60 %, aproximadamente 70 %, aproximadamente 75 %, aproximadamente 80 %, aproximadamente 81 %, aproximadamente 82 %, aproximadamente 83 %, aproximadamente 84 %, aproximadamente 85 %, aproximadamente 86 %, aproximadamente 87 %, aproximadamente 88 %, aproximadamente 89 %, aproximadamente 90 %, aproximadamente 91 %, aproximadamente 92 %, aproximadamente 93 %, aproximadamente 94 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, aproximadamente 99 % o más. En realizaciones particulares, el grado de complementariedad entre una secuencia de repetición de CRISPR y su correspondiente secuencia antirrepetición de ARNtrat, cuando se alinean de manera óptima usando un algoritmo de alineación adecuado, es del 50 %, el 60 %, el 70 %, el 75 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más.

En diferentes realizaciones, el ARNtrat completo puede comprender de aproximadamente 60 nucleótidos a más de aproximadamente 210 nucleótidos. Por ejemplo, el ARNtrat puede ser de aproximadamente 60, aproximadamente 65, aproximadamente 70, aproximadamente 75, aproximadamente 80, aproximadamente 85, aproximadamente 90, aproximadamente 95, aproximadamente 100, aproximadamente 105, aproximadamente 110, aproximadamente 115, aproximadamente 120, aproximadamente 125, aproximadamente 130, aproximadamente 135, aproximadamente 140, aproximadamente 150, aproximadamente 160, aproximadamente 170, aproximadamente 180, aproximadamente 190, aproximadamente 200, aproximadamente 210 o más nucleótidos de longitud. En realizaciones particulares, el ARNtrat tiene 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 150, 160, 170, 180, 190, 200, 210 o más nucleótidos o más de longitud. En realizaciones particulares, el ARNtrat es de aproximadamente 80 a aproximadamente 90 nucleótidos de longitud, incluyendo aproximadamente 80, aproximadamente 81, aproximadamente 82, aproximadamente 83, aproximadamente 84, aproximadamente 85, aproximadamente 86, aproximadamente 87, aproximadamente 88, aproximadamente 89 y aproximadamente 90 de nucleótidos de longitud. En realizaciones particulares, el ARNtrat tiene de 80 a 90 nucleótidos de longitud, incluyendo 80, 81, 82, 83, 84, 85, 86, 87, 88, 89 y 90 nucleótidos de longitud.

En realizaciones particulares, el ARNtrat comprende la secuencia de nucleótidos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125 o una variante o fragmento activo de la misma que cuando está comprendido dentro de un ARN de guía es capaz de dirigir la unión específica de secuencia de una nucleasa guiada por ARN asociada a una secuencia diana de interés. En determinadas realizaciones, una variante de secuencia de ARNtrat activa de una secuencia de tipo silvestre comprende una secuencia de nucleótidos que tiene al menos el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 85 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia con la secuencia de nucleótidos expuesta como las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119 o 125. En determinadas realizaciones, un fragmento de secuencia de ARNtrat activo de una secuencia de tipo silvestre comprende al menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 o más nucleótidos contiguos de la secuencia de nucleótidos expuesta como las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119 o 125.

Se puede considerar que dos secuencias de polinucleótido son sustancialmente complementarias cuando las dos secuencias hibridan entre sí en condiciones rigurosas. De manera análoga, se considera que una NGA se une a una secuencia diana particular de una manera específica de secuencia si el ARN de guía unido a la NGA se une a la secuencia diana en condiciones rigurosas. Por "condiciones rigurosas" o "condiciones de hibridación estrictas" se entienden condiciones en las que las dos secuencias de polinucleótido hibridarán entre sí en un grado detectable mayor que con otras secuencias (por ejemplo, al menos 2 veces con respecto al fondo). Las condiciones rigurosas dependen de la secuencia y serán distintas en circunstancias distintas. Típicamente, las condiciones rigurosas serán las que la concentración de sal es menor de aproximadamente 1,5 M de ion Na, normalmente una concentración de iones Na (u otras sales) de 0,01 a 1,0 M a pH de 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30 °C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos aproximadamente 60 °C para sondas largas (por ejemplo, más de 50 nucleótidos). Las condiciones rigurosas también se pueden conseguir con la adición de agentes desestabilizantes, tales como formamida. Las condiciones de baja rigurosidad ilustrativas incluyen hibridación con una solución tampón de formamida del 30 al 35 %, NaCl 1 M, SDS al 1 % (dodecilsulfato de sodio) a 37 °C y un lavado en SSC de 1X a 2X (SSC 20X = NaCl 3,0 M/citrato trisódico 0,3 M) a de 50 a 55 °C. Las condiciones de rigurosidad moderada ilustrativas incluyen hibridación en formamida del 40 al 45 %, NaCl 1,0 M, SDS al 1 % a 37 °C y un lavado en SSC de 0,5X a 1X a de 55 a 60 °C. Las condiciones de alta rigurosidad ilustrativas incluyen hibridación en formamida al 50 %, NaCl 1 M, SDS al 1 % a 37 °C y un lavado a SSC 0,1X de 60 a 65 °C. Opcionalmente, los tampones de lavado pueden comprender SDS de aproximadamente el 0,1 % a aproximadamente el 1 %. La duración de la hibridación es generalmente menor de aproximadamente 24 horas, normalmente de aproximadamente 4 a aproximadamente 12 horas. La duración del tiempo de lavado será al menos el tiempo suficiente para alcanzar el equilibrio.

La T_m es la temperatura (en condiciones de fuerza iónica y pH definidas) a la cual hibrida un 50 % de una secuencia diana complementaria con una secuencia perfectamente emparejada. Para los híbridos de ADN-ADN, puede calcularse de manera aproximada la T_m mediante la ecuación de Meinkoth y Wahl (1984) Anal. Biochem. 138:267-

284: $T_m = 81,5\text{ }^{\circ}\text{C} + 16,6 (\log \text{ de } M) + 0,41 (\% \text{ de GC}) - 0,61 (\% \text{ de form.}) - 500/L$; donde M es la molaridad de los cationes monovalentes, % de GC es el porcentaje de nucleótidos de guanosina y citosina en el ADN, % de form. es el porcentaje de formamida en la solución de hibridación y L es la longitud del híbrido en pares de bases. En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5 °C más bajas que el punto de fusión térmico (T_m) para la secuencia específica y su complementaria en condiciones de fuerza iónica y pH definidas. Sin embargo, las condiciones altamente rigurosas pueden utilizar una hibridación y/o lavado a 1, 2, 3 o 4 °C más bajos que el punto de fusión térmico (T_m); las condiciones moderadamente rigurosas pueden utilizar una hibridación y/o lavado a 6, 7, 8, 9 o 10 °C más bajos que punto de fusión térmico (T_m); las condiciones de baja rigurosidad pueden utilizar una hibridación y/o lavado a 11, 12, 13, 14, 15 o 20 °C más bajos que el punto de fusión térmico (T_m). Usando la ecuación, las composiciones de hibridación y lavado y la T_m deseada, los expertos habituales en la materia entenderán que se describen inherentemente variaciones en la rigurosidad de las soluciones de hibridación y/o lavado. Puede encontrarse una extensa guía de hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, Parte I, Capítulo 2 (Elsevier, Nueva York); y Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Capítulo 2 (Greene Publishing y Wiley-Interscience, Nueva York). Véase Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2ª ed., Cold Spring Harbor Laboratory Press, Plainview, Nueva York).

La expresión "secuencia específica" también puede referirse a la unión de una secuencia diana con una frecuencia mayor que la unión a una secuencia de fondo aleatoria.

El ARN de guía puede ser un ARN de guía único o un sistema de ARN de guía doble. Un ARN de guía único comprende el ARNcr y el ARNtrt en una única molécula de ARN, mientras que un sistema de ARN de guía doble comprende un ARNcr y un ARNtrt presentes en dos moléculas de ARN distintas, hibridados entre sí a través de al menos una porción de la secuencia de repetición de CRISPR del ARNcr y al menos una porción del ARNtrt, que puede ser total o parcialmente complementaria a la secuencia de repetición de CRISPR del ARNcr. En algunas de las realizaciones en donde el ARN de guía es un ARN de guía único, el ARNcr y el ARNtrt están separados por una secuencia de nucleótidos enlazadora. En general, la secuencia de nucleótidos enlazadora es una que no incluye bases complementarias para evitar la formación de una estructura secundaria dentro o que comprende nucleótidos de la secuencia de nucleótidos enlazadora. En algunas realizaciones, la secuencia de nucleótidos enlazadora entre el ARNcr y el ARNtrt es de al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12 o más nucleótidos de longitud. En realizaciones particulares, la secuencia de nucleótidos enlazadora de un ARN de guía único tiene al menos 4 nucleótidos de longitud. En determinadas realizaciones, la secuencia de nucleótidos enlazadora es la secuencia de nucleótidos expuesta como la SEQ ID NO: 249.

El ARN de guía única o el ARN de guía doble se puede sintetizar químicamente o mediante transcripción *in vitro*. Los ensayos para determinar la unión específica de secuencia entre una NGA y un ARN de guía se conocen en la técnica e incluyen, pero sin limitación, ensayos de unión *in vitro* entre una NGA expresada y el ARN de guía, que se puede etiquetar con una etiqueta detectable (por ejemplo, biotina) y usarse en un ensayo de detección de precipitación en el que el complejo de ARN de guía:NGA se captura a través de la etiqueta detectable (por ejemplo, con perlas de estreptavidina). Se puede utilizar un ARN de guía de control con una secuencia o estructura no relacionada en el ARN de guía como control negativo para la unión no específica de la NGA al ARN. En determinadas realizaciones, el ARN de guía es la SEQ ID NO: 4, 11, 18, 25, 32, 39, 46, 53, 59, 66, 73, 79, 86, 92, 99, 106, 113, 120 o 126, en donde la secuencia espaciadora puede ser cualquier secuencia y se indica como una secuencia de poli-N.

En determinadas realizaciones, el ARN de guía se puede introducir en una célula diana, orgánulo o embrión no humano como molécula de ARN. El ARN de guía se puede transcribir *in vitro* o sintetizar químicamente. En otras realizaciones, se introduce en la célula, orgánulo o embrión no humano una secuencia de nucleótidos que codifica el ARN de guía. En algunas de estas realizaciones, la secuencia de nucleótidos que codifica el ARN de guía está unida operativamente a un promotor (por ejemplo, un promotor de la ARN polimerasa III). El promotor puede ser un promotor nativo o heterólogo para la secuencia de nucleótidos que codifica el ARN de guía.

En diferentes realizaciones, el ARN de guía se puede introducir en una célula diana, orgánulo o embrión no humano como complejo de ribonucleoproteína, como se describe en el presente documento, en donde el ARN de guía está unido a un polipéptido de nucleasa guiada por ARN.

El ARN de guía dirige una nucleasa asociada por ARN guiada a una secuencia de nucleótidos diana particular de interés mediante la hibridación del ARN de guía con la secuencia de nucleótidos diana. Una secuencia de nucleótidos diana puede comprender ADN, ARN o una combinación de ambos y puede ser monocatenaria o bicatenaria. Una secuencia de nucleótidos diana puede ser ADN genómico (es decir, ADN cromosómico), ADN plasmídico o una molécula de ARN (por ejemplo, ARN mensajero, ARN ribosómico, ARN de transferencia, microARN, ARN de interferencia pequeño). La secuencia de nucleótidos diana puede unirse (y en algunas realizaciones, escindir) mediante una nucleasa guiada por ARN *in vitro* o en una célula. La secuencia cromosómica a la que se dirige la NGA puede ser una secuencia cromosómica nuclear, plastidial o mitocondrial. En algunas realizaciones, la secuencia de nucleótidos diana es única en el genoma diana.

La secuencia de nucleótidos diana es adyacente a un motivo adyacente al protoespaciador (PAM). Un motivo adyacente al protoespaciador está generalmente a de aproximadamente 1 a aproximadamente 10 nucleótidos de la secuencia de nucleótidos diana, incluyendo aproximadamente 1, aproximadamente 2, aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9 o aproximadamente 10 nucleótidos de la secuencia de nucleótidos diana. En realizaciones particulares, una PAM se encuentra entre 1 y 10 nucleótidos de la secuencia de nucleótidos diana, incluyendo 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos de la secuencia de nucleótidos diana. El PAM puede estar a 5' o 3' de la secuencia diana. En algunas realizaciones, el PAM está a 3' de la secuencia diana para las NGA divulgadas en el presente documento. En general, la PAM es una secuencia consenso de aproximadamente 3-4 nucleótidos, pero en realizaciones particulares puede ser de 2, 3, 4, 5, 6, 7, 8, 9 o más nucleótidos de longitud. En diferentes realizaciones, la secuencia PAM reconocida por las NGA divulgadas en el presente documento comprende la secuencia consenso establecida como las SEQ ID NO: 7, 14, 21, 28, 35, 42, 49, 62, 69, 79, 82, 95, 102, 109 o 116.

Una nucleasa guiada por ARN que tiene la SEQ ID NO: 83 o una variante o fragmento activo de la misma se une a una secuencia de nucleótidos diana adyacente a una secuencia PAM expuesta como la SEQ ID NO: 82. En algunas realizaciones, la NGA se une a una secuencia guía que comprende una secuencia de repetición de CRISPR expuesta en las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124, respectivamente, o una variante o fragmento activo de la misma, y una secuencia de ARNtrat expuesta en las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125, respectivamente, o una variante o fragmento activo de la misma. Los sistemas de NGA se describen adicionalmente en los Ejemplos 1-3 y las Tablas 1 y 2 de la presente memoria descriptiva.

Se produjeron variantes de NGA APG05586 (SEQ ID NO: 63) que tienen secuencias de aminoácidos de las SEQ ID NO: 570-579. Las NGA que tienen una cualquiera de las SEQ ID NO: 63 y 570-579 pueden unirse a una secuencia de nucleótidos diana adyacente a una secuencia PAM establecida como SEQ ID NO: 79. Estos sistemas de NGA se describen con más detalle en el Ejemplo 5 de la presente memoria descriptiva.

Es bien conocido en la técnica que la especificidad de la secuencia PAM para una enzima nucleasa determinada se ve afectada por la concentración de la enzima (véase, por ejemplo, Karvelis *et al.* (2015) *Genom Biol* 16:253), lo que puede modificarse alterando el promotor utilizado para expresar la NGA o la cantidad de complejo de ribonucleoproteína suministrado a la célula, orgánulo o embrión no humano.

Al reconocer su correspondiente secuencia PAM, la NGA puede escindir la secuencia de nucleótidos diana en un sitio de escisión específico. Como se usa en el presente documento, un sitio de escisión está formado por dos nucleótidos particulares dentro de una secuencia de nucleótidos diana entre los cuales la secuencia de nucleótidos se escinde mediante una NGA. El sitio de escisión puede comprender el 1^{er} y el 2^o, el 2^o y el 3^{er}, el 3^{er} y el 4^o, el 4^o y el 5^o y el 6^o, 7^o y el 8^o o el 8^o y 9^o nucleótidos del PAM en dirección 5' o 3'. En algunas realizaciones, el sitio de escisión puede estar a más de 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos del PAM en la dirección 5' o 3'. Como las NGA pueden escindir una secuencia de nucleótidos diana dando como resultado extremos escalonados, en algunas realizaciones, el sitio de escisión se define basándose en la distancia de los dos nucleótidos del PAM en la cadena positiva (+) del polinucleótido y la distancia de los dos nucleótidos del PAM en la cadena negativa (-) del polinucleótido.

III. Nucleótidos que codifican nucleasas guiadas por ARN, ARN CRISPR y/o ARNtrat

La presente divulgación proporciona polinucleótidos que comprenden las NGA divulgadas en el presente documento, ARNtrat y/o ARNgu y polinucleótidos que comprenden una secuencia nucleotídica que codifica las nucleasas guiadas por ARN divulgadas en el presente documento, ARN de CRISPR, ARNtrat y/o ARNgu. Los polinucleótidos actualmente divulgados incluyen los que comprenden o codifican una secuencia de repetición de CRISPR que comprende la secuencia de nucleótidos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124 o una variante o fragmento activo de la misma que cuando está comprendido dentro de un ARN de guía es capaz de dirigir la unión específica de secuencia de una nucleasa guiada por ARN asociada a una secuencia diana de interés. También se divulgan polinucleótidos que comprenden o codifican un ARNtrat que comprende la secuencia de nucleótidos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125 o una variante o fragmento activo de la misma que cuando está comprendido dentro de un ARN de guía es capaz de dirigir la unión específica de secuencia de una nucleasa guiada por ARN asociada a una secuencia diana de interés.

De acuerdo con la primera realización de la invención, también se proporcionan polinucleótidos que codifican una nucleasa guiada por ARN que comprende la secuencia de aminoácidos establecida como SEQ ID NO: 83, y fragmentos activos o variantes de la misma que conservan la capacidad de unirse a una secuencia nucleotídica diana de una manera específica de la secuencia guiada por ARN.

No se pretende que el uso del término "polinucleótido" o "molécula de ácido nucleico" limite la presente divulgación a polinucleótidos que comprenden ADN. Los expertos habituales en la materia reconocerán que los polinucleótidos pueden comprender ribonucleótidos (ARN) y combinaciones de ribonucleótidos y desoxirribonucleótidos. Dichos desoxirribonucleótidos y ribonucleótidos incluyen moléculas tanto de origen natural como análogos sintéticos. Estos incluyen ácidos peptidonucleicos (los APN), quimeras de APN-ADN, ácidos nucleicos bloqueados (los LNA) y

secuencias unidas a fosfotiorato. Los polinucleótidos divulgados en el presente documento también abarcan todas las formas de secuencias que incluyen, aunque no de forma limitativa, formas monocatenarias, formas bicatenarias, híbridos de ADN-ARN, estructuras triples, estructuras de tallo y bucle, y similares.

- 5 Se pueden optimizar los codones de las moléculas de ácido nucleico que codifican las NGA para su expresión en un organismo de interés. Una secuencia codificante "con codones optimizados" es una secuencia codificante polinucleotídica que tiene su frecuencia de uso de codones diseñada para imitar la frecuencia de uso de codones preferida o las condiciones de transcripción de una célula hospedadora particular. La expresión en la célula hospedadora u organismo particular se potencia como resultado de la alteración de uno o más codones a nivel de
- 10 ácido nucleico de manera que la secuencia de aminoácidos traducida no cambia. Se pueden optimizar los codones de las moléculas de ácido nucleico, ya sea total o parcialmente. En la técnica están disponibles tablas de codones y otras referencias que proporcionan información de preferencia para una amplia gama de organismos (véase, por ejemplo, Campbell y Gowri (1990) *Plant Physiol.* 92:1-11 en cuanto a un análisis del uso de codones preferido para plantas). Existen métodos en la técnica para sintetizar genes preferentes de plantas o secuencias codificantes optimizadas para codones de mamíferos (por ejemplo, humanas). Véase, por ejemplo, las patentes de Estados Unidos n.º 5.380.831 y 5.436.391 y Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498.

- Los polinucleótidos que codifican las NGA, los ARNcr, los ARNtrat y/o los ARNngu proporcionados en el presente documento se pueden proporcionar en casetes de expresión para la expresión *in vitro* o la expresión en una célula, orgánulo, embrión no humano u organismo de interés. El casete incluirá secuencias reguladoras 5' y 3' unidas operativamente a un polinucleótido que codifica una NGA, ARNcr, los ARNtrat y/o los ARNngu proporcionados en el presente documento que permitan la expresión del polinucleótido. El casete puede contener además al menos el gen adicional que se va a cotransformar en el organismo. Cuando se incluyen genes o elementos adicionales, los
- 20 componentes están unidos operativamente. La expresión "unido operativamente" se entiende que significa un enlace funcional entre dos o más elementos. Por ejemplo, un enlace operable entre un promotor y una región codificante de interés (por ejemplo, región que codifica una NGA, ARNcr, los ARNtrat y/o los ARNngu) es un enlace funcional que permite la expresión de la región codificante de interés. Los elementos unidos operativamente pueden ser contiguos o no contiguos. Cuando se utiliza para hacer referencia a la unión de dos regiones codificantes de proteína, por unido operativamente se entiende que las regiones codificantes están en el mismo marco de lectura. Como alternativa, el
- 30 gen o genes o el elemento o elementos adicionales pueden proporcionarse en múltiples casetes de expresión. Por ejemplo, la secuencia de nucleótidos que codifica una NGA actualmente divulgada puede estar presente en un casete de expresión, mientras que la secuencia de nucleótidos que codifica un ARNcr, ARNtrat o el ARN de guía completo puede estar en un casete de expresión distinto. Dicho casete de expresión está provisto de una pluralidad de sitios de restricción y/o sitios de recombinación para la inserción de los polinucleótidos que estarán bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente un gen marcador de selección.

- El casete de expresión incluirá en la dirección de la transcripción de 5'-3', una región de inicio de la transcripción (y, en algunas realizaciones, de la traducción) (es decir, un promotor), un polinucleótido que codifica NGA, ARNcr, ARNtrat y/o ARNsg de acuerdo con aspectos de la invención, y una región de terminación de la transcripción (y en algunas realizaciones, de la traducción) (es decir, región de terminación,) funcional en el organismo de interés. Los promotores son capaces de dirigir o impulsar la expresión de una secuencia codificante en una célula hospedadora. Las regiones reguladoras (por ejemplo, promotores, regiones reguladoras de la transcripción y regiones de terminación de la traducción) pueden ser endógenas o heterólogas para la célula hospedadora o entre sí. Como se usa en el presente
- 45 documento, "heteróloga" en referencia a una secuencia es una secuencia que se origina de una especie ajena o, si es de la misma especie, está sustancialmente modificada con respecto a su forma natural en cuanto a su composición y/o locus genómico mediante intervención humana deliberada. Como se usa en el presente documento, un gen quimérico comprende una secuencia codificante unida operativamente con una región de inicio de la transcripción que es heteróloga para la secuencia codificante.

- Se encuentran disponibles regiones de terminación convenientes del plásmido Ti de *A. tumefaciens*, tales como las regiones de terminación de octopina sintasa y nopalina sintasa. Véase también Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262: 141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5: 141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91: 151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; y Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

- Las señales reguladoras adicionales incluyen, pero sin limitación, sitios de inicio de la transcripción, operadores, activadores, potenciadores, otros elementos reguladores, sitios de unión a ribosomas, un codón de inicio, señales de terminación y similares. Véase, por ejemplo, las patentes de EE. UU. n.º 5.039.523 y 4.853.331; el documento EPO 0480762A2; Sambrook *et al.* (1992) *Molecular Cloning: A Laboratory Manual*, ed. Maniatis *et al.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), en lo sucesivo en el presente documento "Sambrook 11"; Davis *et al.*, eds. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory Press), Cold Spring Harbor, N.Y., y las referencias citadas en el presente documento.

- En la preparación del casete de expresión, los diferentes fragmentos de ADN pueden manipularse, para proporcionar las secuencias de ADN en la orientación apropiada y, según sea apropiado, en la fase de lectura adecuada. Para este

fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN o pueden estar implicadas otras manipulaciones para proporcionar sitios de restricción convenientes, retirada de ADN superfluo, retirada de sitios de restricción, o similares. Para este fin, pueden participar mutagénesis *in vitro*, reparación de cebadores, restricción, hibridación, resustituciones, por ejemplo, transiciones y transversiones.

Puede usarse una serie de promotores en la práctica de los aspectos de la invención. Los promotores pueden seleccionarse basándose en el resultado deseado. Los ácidos nucleicos pueden combinarse con promotores constitutivos, inducibles, específicos de la fase de crecimiento, específicos de tipo celular, preferidos por tejido, específicos de tejido, u otros promotores para la expresión en el organismo de interés. Véase, por ejemplo, los promotores expuestos en el documento WO 99/43838 y en las patentes de EE. UU. n.º: 8,575,425; 7,790,846; 8,147,856; 8,586,832; 7,772,369; 7,534,939; 6,072,050; 5,659,026; 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; y 6.177.611.

Para la expresión en plantas, los promotores constitutivos también incluyen el promotor 35S de CaMV (Odell *et al.* (1985) *Nature* 313:810-812); de actina de arroz (McElroy *et al.* (1990) *Plant Cell* 2:163-171); de ubiquitina (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 y Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); y MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730).

Los ejemplos de promotores inducibles son el promotor Adh1 que es inducible por hipoxia o estrés por frío, el promotor Hsp70 que es inducible por estrés térmico, el promotor PPKK y el promotor de la pepcarboxilasa, ambos inducibles por la luz. También son útiles promotores que son químicamente inducibles, tales como el promotor In2-2 que es inducido por protectores (patente de EE. UU. n.º 5.364.780), el promotor Axig1 que es inducido por auxina y específico de tapete pero también activo en el callo (documento PCT US01/22169), los promotores sensibles a esteroides (véase, por ejemplo, el promotor ERE que es inducido por estrógenos y el promotor inducible por glucocorticoides en Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 y McNellis *et al.* (1998) *Plant J.* 14(2):247-257) y promotores inducibles por tetraciclina y reprimibles por tetraciclina (véase, por ejemplo, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237 y las patentes de EE. UU. N.º 5.814.618 y 5.789.156).

Se pueden utilizar promotores específicos de tejido o preferidos de tejido para dirigir la expresión de una construcción de expresión dentro de un tejido particular. En determinadas realizaciones, los promotores específicos de tejido o preferidos de tejido son activos en el tejido vegetal. Los ejemplos de promotores bajo control del desarrollo en plantas incluyen promotores que inician la transcripción preferentemente en determinados tejidos, tales como hojas, raíces, frutos, semillas o flores. Un promotor "específico de tejido" es un promotor que inicia la transcripción sólo en determinados tejidos. A diferencia de la expresión constitutiva de genes, la expresión específica de tejido es el resultado de varios niveles interactivos de regulación génica. Como tales, puede ser preferible utilizar promotores de especies de plantas homólogas o estrechamente relacionadas para lograr una expresión eficaz y fiable de transgenes en tejidos particulares. En algunas realizaciones, la expresión comprende un promotor preferido de tejido. Un promotor "preferido por el tejido" es un promotor que inicia la transcripción de manera preferente, pero no necesariamente total o exclusivamente en determinados tejidos.

En algunas realizaciones, las moléculas de ácido nucleico que codifican una NGA, ARNcr y/o ARNtrt comprenden un promotor específico del tipo celular. Un promotor "específico de tipo celular" es un promotor que impulsa principalmente la expresión en determinados tipos de células en uno o más órganos. Algunos ejemplos de células vegetales en las que los promotores específicos de tipo celular funcionales en plantas pueden ser principalmente activos incluyen, por ejemplo, células BETL, células vasculares en raíces, hojas, células tallo y células madre. Las moléculas de ácido nucleico también pueden incluir promotores preferidos de tipo celular. Un promotor de "preferido de tipo celular" es un promotor que dirige principalmente la expresión, pero no necesariamente total o exclusivamente, en determinados tipos celulares en uno o más órganos. Algunos ejemplos de células vegetales en las que los promotores preferidos de tipo celular funcionales en plantas pueden ser preferentemente activos incluyen, por ejemplo, células BETL, células vasculares en raíces, hojas, células tallo y células madre.

Las secuencias de ácido nucleico que codifican las NGA, los ARNcr, los ARNtrt y/o los ARNgu pueden unirse operativamente a una secuencia promotora que sea reconocida por una ARN polimerasa de fago, por ejemplo, para la síntesis *in vitro* de ARNm. En dichas realizaciones, el ARN transcrito *in vitro* se puede purificar para su uso en los métodos descritos en el presente documento. Por ejemplo, la secuencia promotora puede ser una secuencia promotora de T7, T3, o SP6 o una variación de una secuencia promotora de T7, T3 o SP6. En dichas realizaciones, la proteína y/o los ARN expresados se pueden purificar para su uso en los métodos de modificación del genoma descritos en el presente documento.

En determinadas realizaciones, el polinucleótido que codifica la NGA, ARNcr, el ARNtrt y/o el ARNgu también pueden unirse a una señal de poliadenilación (por ejemplo, señal poliA de SV40 y otras señales funcionales en plantas) y/o al menos una secuencia de terminación de la transcripción. Además, la secuencia que codifica la NGA también puede estar unida a secuencia(s) que codifican al menos una señal de localización nuclear, al menos el dominio de penetración celular y/o al menos el péptido señal capaz de transportar proteínas a ubicaciones subcelulares particulares, como se describe en cualquier otra parte del presente documento.

El polinucleótido que codifica la NGA, ARNcr, el ARNtrat y/o el ARNgu puede estar presentes en un vector o en múltiples vectores. Un "vector" se refiere a una composición polinucleotídica para transferir, suministrar o introducir un ácido nucleico en una célula hospedadora. Algunos vectores adecuados incluyen vectores plasmídicos, fagémidos, cósmidos, artificiales/minicromosomas, transposones y vectores víricos (por ejemplo, vectores lentivirales, vectores virales adenoasociados, vector baculoviral). El vector puede comprender secuencias de control de la expresión adicionales (por ejemplo, secuencias potenciadoras, secuencias Kozak, secuencias de poliadenilación, secuencias de terminación de la transcripción), secuencias de marcadores de selección (por ejemplo, genes de resistencia a antibióticos), orígenes de replicación, y similares. Se puede encontrar información adicional en "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, Nueva York, 2003 o "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3ª edición, 2001.

El vector también puede comprender un gen marcador de selección para la selección de células transformadas. Se utilizan genes marcadores de selección para la selección de células o tejidos transformados. Los genes marcadores incluyen genes que codifican resistencia a antibióticos, tales como los que codifican la neomicina fosfotransferasa II (NEO) y la higromicina fosfotransferasa (HPT), así como genes que confieren resistencia a compuestos herbicidas, tales como glufosinato de amonio, bromoxinilo, imidazolinonas y 2,4-diclorofenoxiacetato (2,4-D).

En algunas realizaciones, el casete o vector de expresión que comprende la secuencia que codifica el polipéptido NGA puede comprender además una secuencia que codifica un ARNcr y/o un ARNtrat, o el ARNcr y el ARNtrat combinados para crear un ARN guía. La secuencia o secuencias que codifican el ARNcr y/o el ARNtrat pueden unirse operativamente a al menos una secuencia de control transcripcional para la expresión del ARNcr y/o el ARNtrat en el organismo o célula hospedadora de interés. Por ejemplo, el polinucleótido que codifica el ARNcr y/o el ARNtrat puede unirse operativamente a una secuencia promotora que sea reconocida por la ARN polimerasa III (Pol III). Algunos ejemplos de promotores de Pol III adecuados incluyen, pero sin limitación, promotores de ARN U6, U3, H1 y 7SL de mamíferos y promotores U6 y U3 de arroz.

Como se ha indicado, las construcciones de expresión que comprenden secuencias de nucleótidos que codifican las NGA, ARNcr, el ARNtrat y/o el ARNgu se pueden usar para transformar organismos de interés. Los métodos de transformación implican la introducción de una construcción nucleotídica en un organismo de interés. Por "introducir" se entiende introducir la construcción de nucleótidos en la célula hospedadora de tal manera que la construcción obtenga acceso al interior de la célula hospedadora. Los métodos de la invención no requieren un método particular para introducir una construcción nucleotídica en un organismo hospedador, sólo que la construcción nucleotídica obtenga acceso al interior de al menos una célula del organismo hospedador. La célula hospedadora puede ser una célula eucariota o procariota. En realizaciones particulares, la célula hospedadora eucariota es una célula vegetal, una célula de mamífero, una célula de ave o una célula de insecto. En algunas realizaciones, la célula eucariota que comprende o expresa una NGA divulgada en el presente documento o que ha sido modificada por una NGA divulgada en el presente documento es una célula humana. En algunas realizaciones, la célula eucariota que comprende o expresa una NGA divulgada en el presente documento o que ha sido modificada por una NGA divulgada en el presente documento es una célula de origen hematopoyético, tal como una célula inmunitaria (es decir, una célula del sistema inmunitario innato o adaptativo), incluida, entre otras, un linfocito B, un linfocito T, un linfocito citotóxico natural (NK), una célula madre pluripotente, una célula madre pluripotente inducida, un linfocito T receptor de antígeno quimérico (CAR-T), un monocito, un macrófago y una célula dendrítica.

Se conocen en la técnica métodos para introducir construcciones nucleotídicas en plantas y otras células hospedadoras, que incluyen, aunque no de forma limitativa, métodos de transformación estable, métodos de transformación transitoria y métodos mediados por virus.

Los métodos dan como resultado un organismo transformado, tal como una planta, incluyendo plantas completas, así como órganos de plantas (por ejemplo, hojas, tallos, raíces, etc.), semillas, células vegetales, propágulos, embriones vegetales y la descendencia de las mismas. Las células vegetales pueden estar diferenciadas o no diferenciadas (por ejemplo, callos, células de cultivo en suspensión, protoplastos, células foliares, células de la raíz, células del floema, polen).

"Organismos transgénicos" u "organismos transformados" u organismos o células o tejidos "transformados de forma estable" se refiere a organismos que han incorporado o integrado un polinucleótido que codifica una NGA, el ARNcr y/o el ARNtrat de conformidad con los aspectos de la invención. Se reconoce que también pueden incorporarse a la célula hospedadora otras secuencias de ácidos nucleicos o fragmentos de ADN exógenos o endógenos. La transformación mediada por *Agrobacterium* y por biolística siguen siendo los dos enfoques predominantemente empleados para la transformación de células vegetales. Sin embargo, la transformación de una célula hospedadora puede realizarse mediante infección, transfección, microinyección, electroporación, microproyección, biolística o bombardeo de partículas, electroporación, fibras de sílice/carbono, mediada por ultrasonido, mediada por PEG, coprecipitación con fosfato de calcio, técnica de polimerización DMSO, procedimiento de DEAE dextrano y mediada por virus, mediada por liposomas y similares. La introducción mediada por virus de un polinucleótido que codifica una NGA, ARNcr y/o ARNtrat incluye introducción y expresión mediada por retrovirus, lentivirus, adenovirus y adenoasociados, así como el uso de Caulimovirus, Geminivirus y virus ARN de plantas.

Los protocolos de transformación, así como los protocolos para introducir polipéptidos o secuencias de polinucleótido en plantas, pueden variar dependiendo del tipo de célula hospedadora (por ejemplo, célula vegetal monocotiledónea o dicotiledónea) diana de la transformación. Los métodos de transformación son conocidos en la técnica e incluyen los expuestos en las patentes de Estados Unidos N.º: 8,575,425; 7,692,068; 8,802,934; 7,541,517. Véase, también, Rakoczy-Trojanowska, M. (2002) *Cell Mol Biol Lett.* 7:849-858; Jones *et al.* (2005) *Plant Methods* 1:5; Rivera *et al.* (2012) *Physics of Life Reviews* 9:308-345; Bartlett *et al.* (2008) *Plant Methods* 4:1-12; Bates, G.W. (1999) *Methods in Molecular Biology* 111:359-366; Binns y Thomashow (1988) *Annual Reviews in Microbiology* 42:575-606; Christou, P. (1992) *The Plant Journal* 2:275-281; Christou, P. (1995) *Euphytica* 85:13-27; Tzfira *et al.* (2004) *TRENDS in Genetics* 20:375-383; Yao *et al.* (2006) *Journal of Experimental Botany* 57:3737-3746; Zupan y Zambryski (1995) *Plant Physiology* 107:1041-1047; Jones *et al.* (2005) *Plant Methods* 1:5;

La transformación puede dar como resultado una incorporación estable o transitoria del ácido nucleico en la célula. Por "transformación estable" se entiende que la construcción nucleotídica introducida en una célula hospedadora se integra en el genoma de la célula hospedadora y es capaz de ser heredada por su progenie. Por "transformación transitoria" se entiende que un polinucleótido se introduce en la célula hospedadora y no se integra en su genoma.

Se conocen en la técnica métodos para la transformación de cloroplastos. Véase, por ejemplo, Svab *et al.* (1990) *Proc. Natl. Acad. Sci. EE.UU.* 87:8526-8530; Svab y Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917; Svab y Maliga (1993) *EMBO J.* 12:601-606. El método se basa en la administración con una pistola de partículas de ADN que contiene un marcador de selección y en el direccionamiento del ADN al genoma plástmico mediante recombinación homóloga. Además, la transformación de plástmicos puede lograrse mediante la transactivación de un transgén silencioso portado el plástmico mediante la expresión preferida de tejido de una ARN polimerasa codificada en el núcleo y dirigida al plástmico. Tal sistema se ha publicado en McBride *et al.* (1994) *Proc. Nacional. Acad. Ciencia. EE.UU* 91:7301-7305.

Las células que han sido transformadas pueden convertirse en un organismo transgénico, tal como una planta, en conformidad con las formas convencionales. Véase, por ejemplo, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. Estas plantas pueden después cultivarse y polinizarse con la misma cepa transformada o diferentes cepas, y el híbrido resultante que tiene expresión constitutiva de la característica fenotípica deseada. Pueden cultivarse dos o más generaciones para garantizar que la expresión de la característica fenotípica deseada se mantenga de forma estable y se herede, y después se recolectan semillas para garantizar que se ha conseguido la expresión de la característica fenotípica deseada. De esta manera, la presente invención permite semillas transformadas (también denominadas "semillas transgénicas") que tienen una construcción nucleotídica de acuerdo con aspectos de la invención, por ejemplo, un casete de expresión, incorporado de forma estable en su genoma.

Como alternativa, las células que se han transformado pueden introducirse en un organismo. Estas células podrían haberse originado en el organismo, en donde las células se transforman en un enfoque *ex vivo*.

Las secuencias proporcionadas en el presente documento pueden usarse para la transformación de cualquier especie vegetal, incluyendo, aunque no de forma limitativa, monocotiledóneas o dicotiledóneas. Los ejemplos de plantas de interés incluyen, pero sin limitación, gramíneas (maíz), sorgo, trigo, girasol, tomate, crucíferas, pimientos, patata, algodón, arroz, soja, remolacha azucarera, caña de azúcar, tabaco, cebada y colza, *Brassica* sp., alfalfa, centeno, mijo, cártamo, cacahuetes, batata, mandioca, café, coco, piña, árboles de cítricos, cacao, té, plátano, aguacate, higuera, guayaba, mango, aceituna, papaya, anacardo, macadamia, almendra, avena, verduras, plantas ornamentales y coníferas.

Las verduras incluyen, pero sin limitación, tomates, lechuga, judías verdes, habas de Lima, guisantes y miembros del género *Curcumis* tales como pepino, melón cantalupo y melón. Las plantas ornamentales incluyen, pero sin limitación, azalea, hortensia, hibisco, rosas, tulipán, narcisos, petunias, clavel, flor de pascua y crisantemo. Preferentemente, las plantas de la presente invención son plantas de cultivo (por ejemplo, maíz, sorgo, trigo, girasol, tomate, crucíferas, pimientos, patata, algodón, arroz, soja, remolacha azucarera, caña de azúcar, tabaco, cebada, colza, etc.).

Como se usa en el presente documento, el término planta incluye células vegetales, protoplastos vegetales, cultivos de tejido de células vegetales a partir de los que pueden regenerarse plantas, callos de plantas, agrupaciones de plantas y células vegetales que están intactos en plantas o partes de plantas tales como embriones vegetales, polen, óvulos, semillas, hojas, flores, ramas, frutos, nueces, mazorcas, elote, vainas, pedúnculos, raíces, puntas de raíces, anteras y similares. Se entiende que grano significa la semilla madura producida por cultivadores comerciales para fines distintos del cultivo o la reproducción de las especies. Aspectos de la invención también facilitan la progenie, variantes y mutantes de las plantas regeneradas, siempre que estas partes comprendan los polinucleótidos introducidos. Además se proporciona un producto o subproducto vegetal procesado que conserva las secuencias divulgadas en el presente documento, incluyendo, por ejemplo, harina de soja.

Los polinucleótidos que codifican las NGA, los ARNcr y/o los ARNtrat también se pueden usar para transformar cualquier especie procariota, incluyendo, pero sin limitación, arqueas y bacterias (por ejemplo, *Bacillus* sp., *Klebsiella* sp., *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., *Agrobacterium*, *Lactobacillus* sp.).

Los polinucleótidos que codifican las NGA, los ARNcr y/o los ARNtrat se pueden usar para transformar cualquier especie eucariota, que incluyen, aunque no de forma limitativa, animales (por ejemplo, mamíferos, insectos, peces, aves y reptiles), hongos, amebas, algas y levaduras.

Se pueden usar métodos convencionales de transferencia de genes víricos o no víricos para introducir ácidos nucleicos en células o tejidos diana de mamíferos, insectos o aves. Dichos métodos se pueden usar para administrar ácidos nucleicos que codifican componentes de un sistema de NNA en células en cultivo, o en un organismo hospedador. Los sistemas de administración de vectores no víricos incluyen plásmidos de ADN, ARN (por ejemplo, un transcrito de un vector descrito en el presente documento), ácido nucleico desnudo y ácido nucleico en complejo con un vehículo de administración, tal como un liposoma. Los sistemas de administración de vectores víricos incluyen virus de ADN y ARN, que tienen genomas episómicos o integrados después de la administración a la célula. Para una revisión de los procedimientos de terapia génica, véase Anderson, *Science* 256: 808- 813 (1992); Nabel y Feigner, *TIBTECH* 11:211-217 (1993); Mitani y Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10): 1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer y Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, en *Current Topics in Microbiology and Immunology*, Doerfler y Bohm (eds) (1995); y Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

Los métodos de administración no vírica de ácidos nucleicos incluyen lipofección, nucleofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, policación o lípido: conjugados de ácido nucleico, ADN desnudo, viriones artificiales y captación de ADN potenciada por un agente. La lipofección se describe en, por ejemplo, las Patentes de los Estados Unidos n.º 5.049.386, 4.946.787; y 4.897.355) y los reactivos de lipofección se comercializan en el mercado (por ejemplo, Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros adecuados para la lipofección eficaz de polinucleótidos con reconocimiento de receptores incluyen los de Feigner, documento WO 91/17424; documento WO 91/16024. La administración puede ser a células (por ejemplo, administración *in vitro* o *ex vivo*) o a tejidos diana (por ejemplo, administración *in vivo*). La preparación de los complejos de lípido:ácido nucleico, incluyendo los liposomas diana tales como complejos de inmunolípidos, es muy conocida por el experto en la materia (véase, por ejemplo, Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291- 297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); patentes de Estados Unidos números 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028, y 4.946.787).

El uso de sistemas basados en ARN o ADN vírico para la administración de ácidos nucleicos aprovecha procesos altamente evolucionados para el direccionamiento de un virus a células específicas del organismo y el transporte de la carga vírica al núcleo. Los vectores virales se pueden administrar directamente a los pacientes (*in vivo*) o pueden usarse para tratar células *in vitro* y las células modificadas pueden administrarse opcionalmente a pacientes (*ex vivo*). Los sistemas basados en virus convencionales podrían incluir vectores de retrovirus, lentivirus, adenovirus, adenoasociados y del virus herpes simple para la transferencia génica. La integración en el genoma del hospedador es posible con los métodos de transferencia génica de retrovirus, lentivirus, y virus adenoasociados, lo que a menudo tiene como resultado la expresión a largo plazo del transgén insertado. Además, se han observado altas eficiencias de transducción en muchos tipos diferentes celulares y tejidos diana.

El tropismo de un retrovirus puede alterarse incorporando proteínas de la envoltura extrañas, expandiendo la población diana potencial de células diana. Los vectores lentivíricos son vectores retrovíricos que son capaces de transducir o infectar células que no se dividen y normalmente producen títulos víricos altos. La selección de un sistema de transferencia génica retrovírico dependería por lo tanto del tejido diana. Los vectores retrovíricos están compuestos por repeticiones terminales largas que actúan en *cis* con capacidad de empaquetamiento de hasta 6-10 kb de secuencia extraña. Las LTR que actúan en *cis* mínimas son suficientes para la replicación y el empaquetamiento de los vectores, que se usan después para integrar el gen terapéutico en la célula diana para proporcionar una expresión transgénica permanente. Los vectores retrovíricos ampliamente utilizados incluyen los basados en el virus de la leucemia murina (MuLV), virus de la leucemia del gibbon (GaLV), el virus de la inmunodeficiencia de simios (SIV, por sus siglas en inglés), virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véase, por ejemplo, Buchscher *et al.*, *J. Viral.* 66:2731-2739 (1992); Johann *et al.*, *J. Viral.* 66: 1635-1640 (1992); Sommerfelt *et al.*, *Viral.* 176:58-59 (1990); Wilson *et al.*, *J. Viral.* 63:2374-2378 (1989); Miller *et al.*, 1. *Viral.* 65:2220-2224 (1991); documento PCT/US94/05700).

En aplicaciones en las que se prefiere la expresión transitoria, pueden usarse sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una eficacia de transducción muy alta en muchos tipos celulares y no requieren división celular. Con dichos vectores, se han obtenido un alto título y altos niveles de expresión. Este vector puede producirse en grandes cantidades en un sistema relativamente simple. Los vectores de virus adenoasociados ("AAV") también pueden usarse para transducir células con ácidos nucleicos diana, por ejemplo, en la producción *in vitro* de ácidos nucleicos y péptidos, y para procedimientos de terapia génica *in vivo* y *ex vivo* (véase, por ejemplo, West *et al.*, *Virology* 160:38-47 (1987); patente de EE. UU. n.º 4.797.368; documento WO 93/24641; Katin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, 1. *Clin. Invest.* 94:1351 (1994). La construcción de vectores de AAV recombinantes se describe en varias publicaciones, incluyendo la patente de Estados Unidos N.º 5.173.414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat y

Muzyczka, PNAS 81:6466-6470 (1984); y Samulski *et al.*, 1. Viral. 63:03822-3828 (1989). Las células de empaquetamiento se usan normalmente para formar partículas de virus que son capaces de infectar una célula hospedadora. Dichas células incluyen células 293, que empaquetan adenovirus, y células ψ J2 o células PA317, que empaquetan retrovirus.

Los vectores víricos usados en la terapia génica habitualmente se generan produciendo una línea celular que empaqueta un vector de ácido nucleico en una partícula vírica. Los vectores pueden contener normalmente las secuencias víricas mínimas requeridas para el empaquetamiento y la posterior integración en un hospedador, otras secuencias víricas se sustituyen por un casete de expresión para el polinucleótido o polinucleótidos que se desean expresar. Las funciones víricas que faltan se suministran normalmente en *trans* por la línea celular de empaquetamiento. Por ejemplo, Los vectores AAV utilizados en terapia génica normalmente sólo poseen secuencias ITR del genoma del AAV que son necesarias para el empaquetamiento y la integración en el genoma hospedador. El ADN vírico se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes de AAV, concretamente rep y cap, pero que carece de las secuencias ITR.

La línea celular también se puede infectar con adenovirus como un auxiliar. El virus auxiliar propicia la replicación del vector de AAV y la expresión de genes de AAV a partir del plásmido auxiliar. El plásmido auxiliar no está empaquetado en cantidades significativas debido a la falta de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, por ejemplo, tratamiento térmico al que el adenovirus es más sensible que el AAV. Los expertos en la materia conocen métodos adicionales para la administración de ácidos nucleicos a las células. Véase, por ejemplo, US20030087817.

En algunas realizaciones, una célula hospedadora se transfecta de forma transitoria o no transitoria con uno o más vectores descritos en el presente documento. En algunas realizaciones, una célula se transfecta como se produce de forma natural en un sujeto. En algunas realizaciones, se toma una célula transfectada de un sujeto. En algunas realizaciones, la célula procede de células extraídas de un sujeto, tal como una estirpe celular. En algunas realizaciones, la línea celular puede ser de células de mamífero, de insectos o aves. Se conoce una amplia diversidad de líneas celulares para el cultivo de tejidos en la materia. Los ejemplos de líneas celulares incluyen, pero sin limitación, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panel, PC-3, TF1, CTLL-2, CIR, Rat6, CV1, RPTE, A10, T24, 182, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI- 231, HB56, TIB55, lurkat, 145.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4. COS, COS-1, COS-6, COS-M6A, BS-C-1 epitelial de riñón de mono, BALB/3T3 de fibroblasto de embrión de ratón, 3T3 Swiss, 3T3-L1, 132-d5 de fibroblastos fetales humanos; 10.1 de fibroblastos de ratón, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-I, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr-/-, COR-L23, COR-L23/CPR, COR-L235010, CORL23/R23, COS-7, COV-434, CML TI, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, lurkat, células IY, células K562, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/ 0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, líneas celulares OPCN/OPCT, Peer, PNT-1A/ PNT 2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SkBr3, T2, T-47D, T84, línea celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, YAR, y variantes transgénicas de las mismas. Las líneas celulares están disponibles en una variedad de fuentes conocidas por los expertos en la materia (véase, por ejemplo, la Colección Americana de Cultivos Tipo (ATCC) (Manassas, Va.)).

En algunas realizaciones, una célula transfectada con uno o más vectores descritos en el presente documento se usa para establecer una nueva línea celular que comprende una o más secuencias procedentes de vectores. En algunas realizaciones, una célula transfectada transitoriamente con los componentes de un sistema de NGA como se describe en el presente documento (tal como mediante transfección transitoria de uno o más vectores, o transfección con ARN), y modificada a través de la actividad de un sistema de NGA, se usa para establecer una nueva estirpe celular que comprende células que contienen la modificación pero que carecen de cualquier otra secuencia exógena. En algunas realizaciones, las células transfectadas de forma transitoria o no transitoria con uno o más vectores descritos en el presente documento, o las líneas celulares procedentes de tales células se usan para evaluar uno o más compuestos de prueba.

En algunas realizaciones, uno o más vectores descritos en el presente documento se usan para producir un animal transgénico no humano o una planta transgénica. En algunas realizaciones, el animal transgénico es un mamífero, tal como un ratón, rata, hámster, conejo, vaca o cerdo. En algunas realizaciones, el animal transgénico es un ave, tal como un pollo o un pato. En algunas realizaciones, el animal transgénico es un insecto, tal como un mosquito o una garrapata.

IV. Variantes y fragmentos de polipéptidos y polinucleótidos

La presente divulgación proporciona, de acuerdo con aspectos de la invención, variantes activas y fragmentos de una nucleasa guiada por ARN de origen natural (es decir, de tipo salvaje), cuya secuencia de aminoácidos se expone

como la SEQ ID NO: 83, así como variantes y fragmentos activos de repeticiones de CRISPR de origen natural, tal como la secuencia de aminoácidos expuesta como las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124, y variantes activas y fragmentos de ARNtrat de origen natural, tal como la secuencia de aminoácidos expuesta como las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125, y polinucleótidos que codifican los mismos.

Si bien la actividad de una variante o fragmento puede estar alterada en comparación con el polinucleótido o polipéptido de interés, la variante y el fragmento deben conservar la funcionalidad del polinucleótido o polipéptido de interés. Por ejemplo, una variante o fragmento puede una actividad aumentada, una actividad disminuida, un espectro de actividad diferente o cualquier otra alteración en la actividad en comparación con el polinucleótido o polipéptido de interés.

Los fragmentos y variantes de polipéptidos de NGA de origen natural, tal como los divulgados en el presente documento, conservará la actividad de unión al ADN guiada por ARN específica de secuencia. En realizaciones particulares, los fragmentos y variantes de polipéptidos de NGA de origen natural, tal como los divulgados en el presente documento, conservarán la actividad de nucleasa (monocatenaria o bicatenaria).

Los fragmentos y variantes de repeticiones de CRISPR de origen natural, tal como los divulgados en el presente documento, conservarán la capacidad, cuando forman parte de un ARN de guía (que comprende un ARNtrat), de unirse a y guiar una nucleasa guiada por ARN (formando complejo con el ARN de guía) a una secuencia de nucleótidos diana de una manera específica de secuencia.

Los fragmentos y variantes de ARNtrat de origen natural, tal como los divulgados en el presente documento, conservarán la capacidad, cuando forman parte de un ARN de guía (que comprende un ARN de CRISPR), para guiar una nucleasa guiada por ARN (formando complejo con el ARN de guía) hacia una secuencia de nucleótidos diana de una manera específica de secuencia.

El término "fragmento" se refiere a una porción de una secuencia de polinucleótido o polipéptido de la invención. "Fragmentos" o "porciones biológicamente activas" incluyen polinucleótidos que comprenden un número suficiente de nucleótidos contiguos para conservar la actividad biológica (es decir, unirse y dirigir una NGA de una manera específica de secuencia a una secuencia de nucleótidos diana cuando está incluido dentro de un ARN de guía). "Fragmentos" o "porciones biológicamente activas" incluyen polipéptidos que comprenden un número suficiente de restos de aminoácido contiguos para conservar la actividad biológica (es decir, unirse a una secuencia de nucleótidos diana de una manera específica de secuencia cuando forma complejo con un ARN de guía). Los fragmentos de las proteínas NGA incluyen los que son más cortos que las secuencias de longitud completa debido al uso de un sitio de inicio alternativo secuencia abajo. Una porción biológicamente activa de una proteína NGA puede ser un polipéptido que comprende, por ejemplo, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700 o más restos de aminoácidos contiguos de las SEQ ID NO: 1, 8, 15, 22, 29, 36, 43, 50, 56, 63, 70, 76, 83, 89, 96, 103, 110, 117, 123 o 570-579. Dichas porciones biológicamente activas pueden prepararse mediante técnicas recombinantes y evaluarse respecto de su actividad de unión al ADN guiada por ARN específica de secuencia. Un fragmento biológicamente activo de una secuencia de repetición de CRISPR puede comprender al menos 8 aminoácidos contiguos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124. Una porción biológicamente activa de una secuencia de repetición de CRISPR puede ser un polinucleótido que comprende, por ejemplo, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 o 22 nucleótidos contiguos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124. Una porción biológicamente activa de un ARNtrat puede ser un polinucleótido que comprende, por ejemplo, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110 o más nucleótidos contiguos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119 o 125.

En general, "variantes" pretende significar secuencias sustancialmente similares. Para los polinucleótidos, una variante comprende una delección y/o adición de uno o más nucleótidos en uno o más sitios internos en el polinucleótido nativo y/o una sustitución de uno o más nucleótidos en uno o más sitios en el polinucleótido nativo. Como se usa en el presente documento, un polinucleótido o polipéptido "nativo" o "de tipo silvestre" comprende una secuencia de nucleótidos o secuencia de aminoácidos de origen natural, respectivamente. Para los polinucleótidos, las variantes conservadoras incluyen las secuencias que, debido a la degeneración del código genético, codifican la secuencia de aminoácidos nativa del gen de interés. Variantes alélicas de origen natural tales como éstas pueden identificarse con el uso de técnicas de biología molecular muy conocidas, como, por ejemplo, con la reacción en cadena de la polimerasa (PCR) y técnicas de hibridación, como se ha indicado anteriormente. Los polinucleótidos variantes incluyen también polinucleótidos obtenidos sintéticamente, tales como los generados, por ejemplo, mediante el uso de mutagénesis dirigida al sitio pero que todavía codifican el polipéptido o el polinucleótido de interés. En general, las variantes de un polinucleótido particular divulgado en el presente documento tendrán al menos aproximadamente el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 85 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia con el polinucleótido particular como se determina mediante los programas y parámetros de alineamiento de secuencias descritos en otra parte en el presente documento.

Se pueden evaluar también variantes de un polinucleótido particular divulgado en el presente documento (es decir, el polinucleótido de referencia) mediante comparación del porcentaje de identidad de secuencia entre el polipéptido codificado por un polinucleótido variante y el polipéptido codificado por el polinucleótido de referencia. Puede calcularse el porcentaje de identidad entre dos polipéptidos cualesquiera utilizando programas y parámetros de alineamiento de secuencias descritos en otra parte en el presente documento. Cuando se evalúa cualquier par dado de polinucleótidos divulgados en el presente documento por comparación del porcentaje de identidad de secuencia compartido por los dos polipéptidos que codifican, el porcentaje de identidad de secuencia entre los dos polipéptidos codificados es de al menos aproximadamente el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 85 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad de secuencia.

En realizaciones particulares, los polinucleótidos actualmente divulgados codifican un polipéptido de nucleasa guiada por ARN que comprende una secuencia de aminoácidos que tiene al menos el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad con una secuencia de aminoácidos de SEQ ID NO: 83. Una posición aminoacídica de una primera secuencia de aminoácidos "correspondiente a" una posición particular de una segunda secuencia de aminoácidos se refiere a la posición en la primera secuencia de aminoácidos cuando las secuencias de aminoácidos primera y segunda están óptimamente alineadas que se alinea con la posición especificada del resto de aminoácido en la segunda secuencia.

Una variante biológicamente activa de un polipéptido de NGA de acuerdo con aspectos de la invención puede diferir tan solo en aproximadamente 1-15 restos de aminoácido, tan poco como aproximadamente 1-10, tal poco como aproximadamente 6-10, tan poco como 5, tan poco como 4, tan poco como 3, tan solo 2 o tan solo 1 resto de aminoácido. En realizaciones específicas, los polipéptidos pueden comprender un truncamiento aminoterminal o carboxiterminal, que puede comprender al menos una delección de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700 aminoácidos o más del extremo N o C del polipéptido.

En determinadas realizaciones, los polinucleótidos actualmente divulgados comprenden o codifican una repetición de CRISPR que comprende una secuencia de nucleótidos que tiene al menos el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad con la secuencia de nucleótidos expuesta como las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118 o 124.

Los polinucleótidos actualmente divulgados pueden comprender o codificar un ARNtrat que comprende una secuencia de nucleótidos que tiene al menos el 40 %, el 45 %, el 50 %, el 55 %, el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más de identidad con la secuencia de nucleótidos expuesta como las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119 o 125.

Las variantes biológicamente activas de una repetición de CRISPR o ARNtrat pueden diferir en tan sólo aproximadamente 1-15 nucleótidos, tan poco como aproximadamente 1-10, tal poco como aproximadamente 6-10, tan poco como 5, tan poco como 4, tan poco como 3, tan solo 2 o tan solo 1 nucleótido. En realizaciones específicas, los polinucleótidos pueden comprender un truncamiento 5' o 3', que puede comprender al menos una delección de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 95, 100, 105, 110 nucleótidos o más del extremo 5' o 3' del polinucleótido.

Se reconoce que se pueden realizar modificaciones en los polipéptidos de NGA, las repeticiones de CRISPR y los ARNtrat proporcionados en el presente documento creando proteínas y polinucleótidos variantes. Los cambios diseñados por el hombre pueden introducirse mediante la aplicación de técnicas de mutagénesis dirigida. Como alternativa, también pueden identificarse polinucleótidos y/o polipéptidos nativos, aún desconocidos o aún no identificados, estructural y/o funcionalmente relacionados con las secuencias divulgadas en el presente documento, que entren dentro de los aspectos de la presente invención. Se pueden realizar sustituciones conservadoras de aminoácidos en regiones no conservadas que no alteran la función de las proteínas NGA. Como alternativa, pueden realizarse modificaciones que mejoren la actividad de la NGA.

Los polinucleótidos y proteínas variantes abarcan también secuencias y proteínas procedentes de un procedimiento mutagénico y recombinogénico tales como el reordenamiento del ADN. Con dicho procedimiento, una o más proteínas NGA diferentes divulgadas en el presente documento (por ejemplo, las SEQ ID NO: 83) se manipulan para crear una nueva proteína NGA que posea las propiedades deseadas. De esta manera, se generan bibliotecas de polinucleótidos recombinantes a partir de una población de polinucleótidos de secuencias relacionadas que comprenden regiones de secuencias que tienen una identidad de secuencia sustancial y pueden recombinarse de forma homóloga *in vitro* o *in vivo*. Por ejemplo, usando esta estrategia, los motivos de secuencia que codifican un dominio de interés pueden reordenarse entre las secuencias de RGN proporcionadas en el presente documento y otros genes de RGN conocidos.

para obtener un nuevo gen que codifique una proteína con una propiedad mejorada de interés, tal como un aumento de la K_m en el caso de una enzima. Se conocen en la técnica estrategias para dicho reordenamiento de ADN. Véase, por ejemplo, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751; Stemmer (1994) Nature 370:389-391; Cramer *et al.* (1997) Nature Biotech. 15:436-438; Moore *et al.* (1997) J. Mol. Biol. 272:336-347; Zhang *et al.* (1997) Proc. Natl. Acad. Sci. USA 94:4504-4509; Cramer *et al.* (1998) Nature 391:288-291; y las patentes de Estados Unidos n.º 5.605.793 y 5.837.458. Un ácido nucleico "reordenado" es un ácido nucleico producido mediante un procedimiento de reordenamiento tal como cualquier procedimiento de reordenamiento expuesto en el presente documento. Los ácidos nucleicos reordenados se producen recombinando (física o virtualmente) dos o más ácidos nucleicos (o sucesiones de caracteres), por ejemplo, de forma artificial y opcionalmente recursiva. En general, se utilizan una o más etapas de cribado en los procedimientos de reordenamiento para identificar los ácidos nucleicos de interés; esta etapa de cribado se puede realizar antes o después de cualquier etapa de recombinación. En algunas (pero no en todas) realizaciones de reordenamiento, es conveniente realizar múltiples rondas de recombinación antes de la selección para aumentar la diversidad del conjunto que se va a examinar. El proceso global de recombinación y selección se repite opcionalmente de forma recursiva. Dependiendo del contexto, reordenamiento puede referirse a un proceso general de recombinación y selección, o, alternativamente, puede referirse simplemente a las partes recombinatorias del proceso global.

Como se usa en el presente documento, la "identidad de secuencia" o "identidad" en el contexto de dos secuencias de polinucleótido o polipéptido hace referencia a los restos en las dos secuencias que son iguales cuando se alinean para máxima correspondencia a lo largo de una ventana de comparación específica. Cuando el porcentaje de identidad de secuencia se usa en referencia a las proteínas se reconoce que las posiciones de restos que no son idénticas difieren a menudo en sustituciones de aminoácidos conservadoras, donde se sustituyen los restos de aminoácido por otros restos de aminoácido con propiedades químicas similares (p. ej., carga o hidrofobia) y por lo tanto, no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservativas, puede ajustarse por exceso el porcentaje de identidad de secuencia para corregir la naturaleza conservativa de la sustitución. Se dice que las secuencias que difieren en dichas sustituciones conservativas tienen "similitud de secuencia" o "similitud". Los medios para hacer este ajuste son bien conocidos por los expertos en la materia. Típicamente, esto implica puntuar una sustitución conservativa como un emparejamiento erróneo parcial en lugar de completo, aumentando de este modo el porcentaje de identidad de secuencia. Por tanto, por ejemplo, cuando se asigna a un aminoácido idéntico una puntuación de 1 y se asigna a una sustitución no conservadora una puntuación de cero, se asigna a una sustitución conservadora una puntuación de entre cero y 1. La puntuación de sustituciones conservativas se calcula, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

Como se usa en el presente documento, "porcentaje de identidad de secuencia" significa el valor determinado al comparar dos secuencias alineadas de manera óptima a lo largo de una ventana de comparación, en donde la porción de la secuencia polinucleotídica en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las cuales se encuentran la base de ácido nucleico o el resto de aminoácido idénticos en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes entre el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para producir el porcentaje de identidad de secuencia.

A menos que se indique de otro modo, los valores de identidad/similitud de secuencia proporcionados en el presente documento se refieren al valor obtenido usando GAP Versión 10 utilizando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos utilizando una ponderación de GAP de 50 y una ponderación de la longitud de 3 y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud de una secuencia de aminoácidos utilizando una ponderación de GAP de 8 y una ponderación de la longitud de 2 y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de los mismos. Por "programa equivalente" se entiende cualquier programa de comparación de secuencias que, para dos secuencias cualquiera en cuestión, genera un alineamiento que tiene emparejamientos de restos de nucleótidos o aminoácidos idénticos y un porcentaje de identidad de secuencia idéntico en comparación con la alineación correspondiente generada por GAP Versión 10.

Dos secuencias están "alineadas de manera óptima" cuando están alineadas para la puntuación de similitud usando una matriz de sustitución de aminoácidos definida (por ejemplo, BLOSUM62), penalización por existencia de huecos y penalización por extensión de huecos para llegar a la puntuación más alta posible para ese par de secuencias. Las matrices de sustitución de aminoácidos y su uso para cuantificar la similitud entre dos secuencias son bien conocidas en la técnica y se describen, por ejemplo, en Dayhoff *et al.* (1978) "A model of evolutionary change in proteins". En "Atlas of Protein Sequence and Structure", Vol. 5, Supl. 3 (ed. M. O. Dayhoff), pág. 345-352. Natl. Biomed. Res. Found., Washington, D.C. y Henikoff *et al.* (1992) Proc. Natl. Acad. Sci. USA 89: 10915-10919. La matriz BLOSUM62 se utiliza a menudo como matriz de sustitución de puntuación predeterminada en protocolos de alineamiento de secuencias. La penalización por existencia de hueco se impone por la introducción de un hueco de un único aminoácido en una de las secuencias alineadas, y la penalización por extensión de hueco se impone por cada posición de aminoácido vacía adicional insertada en un hueco ya abierto. El alineamiento se define por las posiciones de los aminoácidos de cada secuencia en las que comienza y termina el alineamiento, y opcionalmente por la inserción de un hueco o múltiples huecos en una o ambas secuencias, para llegar a la puntuación más alta posible. Si bien el alineamiento y puntuación

óptimos se pueden lograr manualmente, el procedimiento se facilita mediante el uso de un algoritmo de alineamiento implementado por ordenador, por ejemplo, BLAST 2.0 con huecos, descrito en Altschul *et al.* (1997) Nucleic Acids Res. 25:3389-3402, y puesto a disposición del público en el sitio de internet del National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Los alineamientos óptimos, incluyendo los alineamientos múltiples, se pueden preparar usando, por ejemplo, PSI-BLAST, disponible a través de www.ncbi.nlm.nih.gov y descrito por Altschul *et al.* (1997) Nucleic Acids Res. 25:3389-3402.

Con respecto a una secuencia de aminoácidos que está óptimamente alineada con una secuencia de referencia, un resto de aminoácido "corresponde a" la posición en la secuencia de referencia con la que el resto está emparejado en el alineamiento. La "posición" se indica mediante un número que identifica secuencialmente cada aminoácido en la secuencia de referencia en función de su posición con respecto al extremo aminoterminal. Debido a las delecciones, inserciones, truncamientos, fusiones, etc., que deben tenerse en cuenta a la hora de determinar una alineación óptima, en general, el número de restos de aminoácidos en una secuencia de prueba determinado simplemente contando desde el extremo N terminal no será necesariamente el mismo que el número de su posición correspondiente en la secuencia de referencia. Por ejemplo, en el caso de que haya una delección en una secuencia de prueba alineada, no habrá ningún aminoácido que corresponda a una posición en la secuencia de referencia en el sitio de la delección. Cuando hay una inserción en una secuencia de referencia alineada, esa inserción no corresponderá a ninguna posición de aminoácido en la secuencia de referencia. En el caso de truncamientos o fusiones, puede haber tramos de aminoácidos en la secuencia de referencia o alineada que no corresponden a ningún aminoácido en la secuencia correspondiente.

V. Anticuerpos

También se permiten anticuerpos contra los polipéptidos NGA o las ribonucleoproteínas que comprenden los polipéptidos NGA del cuarto aspecto de la presente invención, incluidos los que tienen la secuencia de aminoácidos establecida como SEQ ID NO: 83 o variantes activas o fragmentos de los mismos. En la técnica se conocen bien los métodos para producir anticuerpos (véase, por ejemplo, Harlow y Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; y la patente de los Estados Unidos n.º 4,196,265). Estos anticuerpos se pueden utilizar en kits para la detección y aislamiento de polipéptidos o ribonucleoproteínas de NGA. Por tanto, la presente divulgación proporciona kits que comprenden anticuerpos que se unen específicamente a los polipéptidos o ribonucleoproteínas descritos en el presente documento, incluyendo, por ejemplo, polipéptidos que tienen la secuencia de SEQ ID NO:83.

VL. Sistemas y complejos de ribonucleoproteína para la unión a una secuencia diana de interés y métodos para producirlos

La presente divulgación proporciona un sistema de acuerdo con el quinto aspecto de la invención para unir una secuencia diana de interés, en donde el sistema comprende al menos un ARN de guía o una secuencia de nucleótidos que codifica el mismo, y al menos una nucleasa guiada por ARN o una secuencia de nucleótidos que codifica la misma. El ARN de guía hibrida con la secuencia diana de interés y también forma un complejo con el polipéptido de NGA, dirigiendo así el polipéptido de NGA para que se una a la secuencia diana. La NGA comprende una secuencia de aminoácidos de SEQ ID NO: 83, o una variante activa o fragmento de la misma de acuerdo con el cuarto aspecto de la invención. En diferentes realizaciones, el ARN de guía comprende una secuencia de repetición de CRISPR que comprende la secuencia de nucleótidos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124, o una variante activa o fragmento del mismo. En realizaciones particulares, el ARN de guía comprende un ARN_{trn} que comprende una secuencia de nucleótidos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125, o una variante activa o fragmento del mismo. El ARN de guía del sistema puede ser un ARN de guía único o un ARN de guía doble. En realizaciones particulares, el sistema comprende una nucleasa guiada por ARN que es heteróloga para el ARN de guía, en donde la NGA y el ARN guía no se encuentran formando complejos uno con otro (es decir, unidos entre sí) en la naturaleza.

El sistema para la unión a una secuencia diana de interés proporcionado en el presente documento puede ser un complejo de ribonucleoproteína, que es al menos una molécula de un ARN unida a al menos una proteína. Los complejos de ribonucleoproteína proporcionados en el presente documento comprenden al menos un ARN de guía y el componente de ARN y una nucleasa guiada por ARN como componente proteico. Dichos complejos de ribonucleoproteína pueden purificarse a partir de una célula u organismo que exprese de forma natural un polipéptido de NGA y haya sido técnicamente diseñado para expresar un ARN de guía particular que sea específico para una secuencia diana de interés. Como alternativa, el complejo ribonucleoproteico puede purificarse a partir de una célula u organismo que haya sido transformada con polinucleótidos que codifican un polipéptido de NGA y un ARN guía y cultivado en condiciones que permiten la expresión del polipéptido de NGA y del ARN guía. Por tanto, se proporcionan métodos para fabricar un polipéptido de NGA o un complejo ribonucleoproteico NGA. Tales métodos comprenden el cultivo de una célula que comprende una secuencia de nucleótidos que codifica un polipéptido NGA y, en algunas realizaciones, una secuencia de nucleótidos que codifica un ARN guía, en condiciones en las que se expresa el polipéptido de NGA (y en algunas realizaciones, el ARN de guía). A continuación, el polipéptido de NGA o la ribonucleoproteína de NGA pueden purificarse a partir de un lisado de las células cultivadas.

Los métodos para purificar un polipéptido NGA o un complejo ribonucleoproteico NGA a partir de un lisado de una muestra biológica son conocidos en la técnica (por ejemplo, exclusión por tamaño y/o cromatografía de afinidad, 2D-PAGE, HPLC, cromatografía de fase inversa, inmunoprecipitación). En métodos particulares, el polipéptido de NGA se produce de forma recombinante y comprende una etiqueta de purificación para facilitar su purificación, incluyendo, pero sin limitación, glutatión-S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tiorredoxina (TRX), poli(NANP), etiqueta de purificación por afinidad en tándem (TAP, por sus siglas en inglés), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, 10xHis, proteína transportadora de biotina carboxilo (BCCP) y calmodulina. En general, el polipéptido de NGA marcado o el complejo ribonucleoproteico de NGA se purifica mediante cromatografía de afinidad por metales inmovilizados. Se apreciará que se pueden usar otros métodos similares conocidos en la técnica, incluyendo otras formas de cromatografía o, por ejemplo, inmunoprecipitación, bien solos o combinados.

Un polipéptido "aislado" o "purificado", o una porción biológicamente activa de los mismos, está sustancial o esencialmente libre de componentes que normalmente acompañan o interactúan con el polipéptido que se encuentran en su entorno de origen natural. Por tanto, un polipéptido aislado o purificado está sustancialmente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente. Una proteína que está sustancialmente libre de material celular incluye preparaciones de proteína que tienen menos de aproximadamente el 30 %, el 20 %, el 10 %, el 5 % o el 1 % (en peso seco) de proteína contaminante. Cuando la proteína del cuarto aspecto de la invención o porción biológicamente activa de la misma se produce recombinantemente, de forma óptima el medio de cultivo representa menos de aproximadamente el 30 %, el 20 %, el 10 %, el 5 % o el 1 % (en peso seco) de precursores químicos o productos químicos que no son proteínas de interés.

Los métodos particulares proporcionados en el presente documento para unirse a y/o escindir una secuencia diana de interés implican el uso de un complejo de ribonucleoproteína de NGA ensamblado *in vitro*. El ensamblaje *in vitro* de un complejo de ribonucleoproteína de NGA se puede realizar usando cualquier método conocido en la técnica en el que un polipéptido de NGA se pone en contacto con un ARN de guía en condiciones que permitan la unión del polipéptido de NGA al ARN de guía. Como se usa en el presente documento, "se pone en contacto", poner en contacto", "puesto en contacto" se refieren a colocar los componentes de una reacción deseada juntos en condiciones adecuadas para llevar a cabo la reacción deseada. El polipéptido de NGA se puede purificar a partir de una muestra biológica, lisado celular o medio de cultivo, producidos mediante traducción *in vitro* o síntesis química. El ARN de guía se puede purificar a partir de una muestra biológica, lisado celular o medio de cultivo, transcribirse *in vitro* o sintetizarse químicamente. El polipéptido de NGA y ARN de guía se pueden poner en contacto en solución (por ejemplo, solución salina tamponada) para permitir el ensamblaje *in vitro* del complejo de ribonucleoproteína de NGA.

VII. Métodos de unión, escisión o modificación de una secuencia diana

La presente divulgación proporciona métodos *in vitro* y *ex vivo* para la unión, escisión y/o modificación de una secuencia de nucleótidos diana de interés de conformidad con aspectos de la invención. Los métodos incluyen suministrar un sistema que comprende al menos un ARN de guía o un polinucleótido que codifica el mismo, y al menos el polipéptido de NGA o un polinucleótido que codifica el mismo a la secuencia diana o una célula, orgánulo o embrión no humano que comprenda la secuencia diana, en donde el método no es un proceso para modificar la identidad de la línea germinal de un ser humano. El NGA comprende la secuencia de aminoácidos de SEQ ID NO: 83, o una variante activa o fragmento de la misma de acuerdo con el cuarto aspecto de la invención. En diferentes realizaciones, el ARN de guía comprende una secuencia de repetición de CRISPR que comprende la secuencia de nucleótidos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124, o una variante activa o fragmento del mismo. En realizaciones particulares, el ARN de guía comprende un ARN_{tr} que comprende la secuencia de nucleótidos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125, o una variante activa o fragmento del mismo. El ARN de guía del sistema puede ser un ARN de guía único o un ARN de guía doble. La NGA del sistema puede ser una NGA sin actividad de nucleasa, tener actividad de nickasa o puede ser un polipéptido de fusión. En algunas realizaciones, el polipéptido de fusión comprende un polipéptido de edición de bases, por ejemplo una citidina desaminasa o una adenosina desaminasa. En otras realizaciones, la proteína de fusión NGA comprende una transcriptasa inversa. En otras realizaciones, la proteína de fusión NGA comprende un polipéptido que recluta miembros de un complejo funcional de reparación de ácidos nucleicos, tal como un miembro de la vía de reparación por escisión de nucleótidos (NER) o de reparación por escisión de nucleótidos acoplada a transcripción (TC-NER) (Wei *et al.*, 2015, PNAS USA 112(27):E3495-504; Troelstra *et al.*, 1992, Cell 71:939-953; Marnef *et al.*, 2017, J Mol Biol 429(9):1277-1288), tal como se describe en la solicitud provisional de Estados Unidos n.º 62/966.203, que se presentó el 27 de enero de 2020, y se incorpora por referencia en su totalidad. En algunas realizaciones, la proteína de fusión NGA comprende CSB (van den Boom *et al.*, 2004, J Cell Biol 166(1):27-36; van Gool *et al.*, 1997, EMBO J 16(19):5955-65; un ejemplo de la cual se establece como la SEQ ID NO: 608), que es un miembro de la vía TC-NER (reparación por escisión de nucleótidos) y funciona en el reclutamiento de otros miembros. En realizaciones adicionales, la proteína de fusión NGA comprende un dominio activo de CSB, tal como el dominio ácido de CSB que comprende restos de aminoácidos 356-394 de la SEQ ID NO: 608 (Teng *et al.*, 2018, Nat Commun 9(1):4115).

En realizaciones particulares, la NGA y/o el ARN de guía son heterólogos para la célula, orgánulo o embrión no

humano en el que se introducen la NGA y/o el ARN de guía (o polinucleótido o polinucleótidos que codifican al menos uno de NGA y ARN de guía).

En las realizaciones en las que el método comprende la administración de un polinucleótido que codifica un ARN guía y/o un polipéptido NGA, la célula o el embrión no humano puede entonces cultivarse en condiciones en las que se expresen el ARN de guía y/o el polipéptido de NGA. En diferentes realizaciones, el método comprende poner en contacto una secuencia diana con un complejo de ribonucleoproteína de NGA. El complejo de ribonucleoproteína de NGA puede comprender una NGA que no tiene actividad de nucleasa o tiene actividad de nickasa. En algunas realizaciones, la NGA del complejo de ribonucleoproteína es un polipéptido de fusión que comprende un polipéptido de edición de bases. En determinadas realizaciones, el método comprende introducir en una célula, orgánulo o embrión no humano que comprende una secuencia diana un complejo de ribonucleoproteína de NGA. El complejo de ribonucleoproteína de NGA puede ser uno que se haya purificado a partir de una muestra biológica, se haya producido de forma recombinante y posteriormente purificado, o se haya ensamblado *in vitro* como se describe en el presente documento. En las realizaciones en donde el complejo de ribonucleoproteína de NGA que se pone en contacto con la secuencia diana o un orgánulo celular, o embrión no humano se ha ensamblado *in vitro*, el método puede comprender además el ensamblaje *in vitro* del complejo antes de la puesta en contacto con la secuencia diana, célula, orgánulo o embrión no humano.

Un complejo de ribonucleoproteína de NGA purificado o ensamblado *in vitro* se puede introducir en una célula, orgánulo o embrión no humano usando cualquier método conocido en la técnica, incluyendo, pero no limitado a, electroporación. Como alternativa, puede introducirse en una célula un polipéptido NGA y/o un polinucleótido que codifique o comprenda el ARN guía, orgánulo o embrión no humano usando cualquier método conocido en la técnica (por ejemplo, electroporación).

Tras la administración o la puesta en contacto con la secuencia diana o célula, orgánulo o embrión no humano que comprenda la secuencia diana, el ARN guía dirige la NGA para que se una a la secuencia diana de forma específica para la secuencia. En las realizaciones en donde la NGA tiene actividad de nucleasa, el polipéptido de NGA escinde la secuencia diana de interés tras la unión. La secuencia diana puede modificarse posteriormente mediante mecanismos de reparación endógenos, tales como unión de extremos no homólogos o reparación dirigida por homología con un polinucleótido donante proporcionado.

Los métodos para medir la unión de un polipéptido de NGA a una secuencia diana se conocen en la técnica e incluyen ensayos de inmunoprecipitación de cromatina, ensayos de cambio de movilidad en gel, ensayos de extracción de ADN, ensayos de indicador, ensayos de captura y detección en microplacas. De manera análoga, los métodos para medir la escisión o modificación de una secuencia diana se conocen en la técnica e incluyen ensayos de escisión *in vitro* o *in vivo* en donde la escisión se confirma mediante PCR, secuenciación o electroforesis en gel, con o sin la unión de un marcador adecuado (por ejemplo, radioisótopo, sustancia fluorescente) a la secuencia diana para facilitar la detección de productos de degradación. Como alternativa, se puede utilizar el ensayo de reacción de amplificación exponencial desencadenada por mellado (NTEXPAR, del inglés *nicking triggered exponential amplification reaction*) (véase, por ejemplo, Zhang *et al.* (2016) Chem. Sci. 7:4951-4957). La escisión *in vivo* se puede evaluar utilizando el ensayo Surveyor (Guschin *et al.* (2010) Methods Mol Biol 649:247-256).

En algunas realizaciones, los métodos implican el uso de un único tipo de NGA formando complejo con más de un ARN de guía. Más de un ARN guía puede dirigirse a diferentes regiones de un único gen o a múltiples genes.

En las realizaciones en donde no se proporciona un polinucleótido donante, una rotura bicatenaria introducida por un polipéptido de NGA se puede reparar mediante un proceso de reparación de unión de extremos no homólogos (NHEJ). Debido a la naturaleza propensa a errores de la NHEJ, la reparación de la rotura bicatenaria puede dar lugar a una modificación de la secuencia diana. Como se usa en el presente documento, una "modificación" en referencia a una molécula de ácido nucleico se refiere a un cambio en la secuencia de nucleótidos de la molécula de ácido nucleico, que puede ser una delección, inserción o sustitución de uno o más nucleótidos, o una combinación de las mismas. La modificación de la secuencia diana puede dar como resultado la expresión de un producto proteico alterado o la inactivación de una secuencia codificante.

En las realizaciones en donde se proporciona un polinucleótido donante, la secuencia donante en el polinucleótido donante puede integrarse o intercambiarse con la secuencia de nucleótidos diana durante el curso de la reparación de la rotura bicatenaria introducida, dando como resultado la introducción de la secuencia donante exógena. Un polinucleótido donante comprende así una secuencia donante que se desea introducir en una secuencia diana de interés. En algunas realizaciones, la secuencia donante altera la secuencia de nucleótidos diana original de modo que la secuencia donante recién integrada no será reconocida ni escindida por la NGA. La integración de la secuencia donante puede mejorarse mediante la inclusión dentro del polinucleótido donante de secuencias flanqueantes, denominados en el presente documento "brazos homólogos" que tienen una identidad de secuencia sustancial con las secuencias que flanquean la secuencia de nucleótidos diana, permitiendo un proceso de reparación dirigido por homología. En algunas realizaciones, los brazos homólogos tienen una longitud de al menos 50 pares de bases, al menos 100 pares de bases y hasta 2000 pares de bases o más, y tienen al menos un 90 %, al menos un 95 % o más, de homología de secuencia con su secuencia correspondiente dentro de la secuencia de nucleótidos diana.

En las realizaciones en donde el polipéptido de NGA introduce roturas bicatenarias escalonadas, el polinucleótido donante puede comprender una secuencia donante flanqueada por protuberancias compatibles, permitiendo la ligación directa de la secuencia donante a la secuencia de nucleótidos diana escindida que comprende protuberancias mediante un proceso de reparación no homólogo durante la reparación de la rotura bicatenaria.

En las realizaciones en donde el método implica el uso de una NGA que es una nickasa (es decir, sólo es capaz de escindir una sola cadena de un polinucleótido bicatenario), el método puede comprender la introducción de dos nickasas NGA que se dirigen a secuencias diana idénticas o solapantes y escinden diferentes cadenas del polinucleótido. Por ejemplo, se puede introducir una nickasa NGA que sólo escinde la cadena positiva (+) de un polinucleótido bicatenario junto con una segunda NGA nickasa que sólo escinde la cadena negativa (-) de un polinucleótido bicatenario.

En diferentes realizaciones, se proporciona un método para la unión a una secuencia de nucleótidos diana y la detección de la secuencia diana, en donde el método comprende introducir en una célula, orgánulo o embrión humano al menos un ARN de guía o un polinucleótido que codifica el mismo, y al menos un polipéptido de NGA o un polinucleótido que codifica el mismo, expresar el ARN de guía y/o el polipéptido de NGA (si se introducen secuencias codificantes), en donde el polipéptido de NGA es una NGA sin actividad de nucleasa y comprende además un marcador de detección, y el método comprende además detectar el marcador de detección. El marcador de detección puede fusionarse a la NGA como una proteína de fusión (por ejemplo, proteína fluorescente) o puede ser una molécula pequeña conjugada a o incorporada dentro del polipéptido de NGA que puede detectarse visualmente o por otros medios.

También se permite en el presente documento métodos para modular la expresión de una secuencia diana o un gen de interés bajo la regulación de una secuencia diana. Los métodos comprenden introducir en una célula, orgánulo o embrión no humano al menos un ARN de guía o un polinucleótido que codifica el mismo, y al menos un polipéptido de NGA o un polinucleótido que codifica el mismo, expresar el ARN de guía y/o el polipéptido de NGA (si se introducen secuencias codificantes), en donde el polipéptido de NGA es una NGA sin actividad de nucleasa. En algunas de estas realizaciones, la NGA sin actividad de nucleasa es una proteína de fusión que comprende un dominio modulador de expresión (es decir, dominio de modificación epigenética, dominio de activación transcripcional o dominio de represión transcripcional) como se describe en el presente documento.

La presente divulgación también proporciona métodos para la unión a y/o la modificación de una secuencia de nucleótidos diana de interés. Los métodos incluyen suministrar un sistema que comprende al menos el ARN de guía o un polinucleótido que codifica el mismo, y al menos el polipéptido de fusión comprende una NGA de la invención y un polipéptido de edición de bases, por ejemplo, una citidina desaminasa o una adenosina desaminasa, o un polinucleótido que codifique el polipéptido de fusión, a la secuencia diana o a una célula, orgánulo o embrión no humano que comprenda la secuencia diana.

Un experto en la materia apreciará que cualquiera de los métodos actualmente divulgados se puede utilizar para dirigirse a una única secuencia diana o a múltiples secuencias diana. Por tanto, los métodos comprenden el uso de un único polipéptido de NGA en combinación con múltiples ARN de guía distintos, que pueden apuntar a múltiples, secuencias distintas dentro de un único gen y/o múltiples genes. También se incluyen en el presente documento métodos en donde se introducen múltiples ARN guía distintos en combinación con múltiples, polipéptidos de NGA distintos. Estos ARN guía y sistemas de polipéptidos de ARN guía/NGA pueden dirigirse a múltiples secuencias diana distintas dentro de un único gen y/o múltiples genes.

En un aspecto, la invención facilita kits que contienen uno cualquiera o más de los elementos divulgados en los métodos y composiciones anteriores. En algunas realizaciones, el kit comprende un sistema de vector e instrucciones para usar el kit. En algunas realizaciones, el sistema de vector comprende (a) un primer elemento regulador unido operativamente a una secuencia de apareamiento de ARNcr y uno o más sitios de inserción para insertar una secuencia de guía en 5' de la secuencia de apareamiento del ARNcr, en donde, cuando se expresa, la secuencia de guía dirige la unión específica de secuencia de un complejo de NGA a una secuencia diana en una célula eucariota, en donde el complejo NGA comprende una enzima NGA en complejo con el polinucleótido de ARN guía; y/o (b) un segundo elemento regulador unido operativamente a una secuencia codificante de una enzima que codifica dicha enzima de NGA que comprende una secuencia de localización nuclear. Los elementos pueden proporcionarse individualmente o en combinaciones, y pueden proporcionarse en cualquier recipiente adecuado, tal como un vial, un frasco o un tubo.

En algunas realizaciones, el kit incluye instrucciones en uno o más idiomas. En algunas realizaciones, un kit comprende uno o más reactivos para su uso en un procedimiento que utiliza uno o más de los elementos descritos en el presente documento. Los reactivos pueden proporcionarse en cualquier recipiente adecuado. Por ejemplo, un kit puede proporcionar uno o más tampones de reacción o almacenamiento. Los reactivos se pueden proporcionar en una forma que sea utilizable en un ensayo particular, o en una forma que precise la adición de uno o más componentes antes de su uso (por ejemplo, en forma concentrada o liofilizada). Un tampón puede ser cualquier tampón, incluyendo, entre otros, un tampón de carbonato de sodio, un tampón de bicarbonato de sodio, un tampón de borato, un tampón

Tris, un tampón MOPS, un tampón HEPES y combinaciones de los mismos. En algunas realizaciones, el tampón es alcalino. En algunas realizaciones, el tampón tiene un pH de aproximadamente 7 a aproximadamente 10.

En algunas realizaciones, el kit comprende uno o más oligonucleótidos correspondientes a una secuencia de guía para su inserción en un vector para unir operativamente la secuencia guía y un elemento regulador. En algunas realizaciones, el kit comprende un polinucleótido molde de recombinación homóloga. En un aspecto, la invención facilita métodos para usar uno o más elementos de un sistema de NGA. El complejo de NGA de la invención facilita un medio eficaz para modificar un polinucleótido diana. El sistema de NGA de la invención facilita una amplia diversidad de utilidades que incluyen la modificación (por ejemplo, delección, inserción, translocación, inactivación, activación, edición de bases) un polinucleótido diana en una multiplicidad de tipos celulares. Como tal, el sistema de NGA de la invención facilita un amplio espectro de aplicaciones en, por ejemplo, terapia génica, cribado de fármacos, diagnóstico y pronóstico de enfermedades. Un sistema de NGA ilustrativo, o complejo de NGA, comprende una enzima de NGA formando complejo con una secuencia de guía hibridada a una secuencia diana dentro del polinucleótido diana.

VIII. Polinucleótidos diana

De acuerdo con aspectos de la invención, la invención proporciona métodos *in vitro* o *ex vivo* para modificar un polinucleótido diana en una célula eucariota, en donde los métodos no son procesos para modificar la identidad genética de la línea germinal de seres humanos. En algunas realizaciones, el método comprende tomar muestras de una célula o población de células de un animal humano o no humano o una planta (incluidas microalgas) y modificar la célula o células. El cultivo puede tener lugar en cualquier fase *ex vivo*. La célula o células pueden incluso reintroducirse en el animal no humano o planta (incluidas microalgas).

Utilizando la variabilidad natural, los fitomejoradores combinan los genes más útiles para obtener las cualidades convenientes, tales como el rendimiento, la calidad, la uniformidad, la rusticidad y la resistencia contra plagas. Estas cualidades convenientes también incluyen el crecimiento, las preferencias de duración del día, los requisitos de temperatura, la fecha de inicio del desarrollo floral o reproductivo, el contenido de ácidos grasos, la resistencia a insectos, resistencia a enfermedades, la resistencia a los nematodos, la resistencia a hongos, la resistencia a herbicidas, la tolerancia a diversos factores ambientales, incluidos sequía, calor, humedad, frío, el viento y las condiciones adversas del suelo, incluida la alta salinidad. Las fuentes de estos genes útiles incluyen variedades nativas o extrañas, variedades tradicionales, plantas silvestres emparentadas y mutaciones inducidas, por ejemplo, el tratamiento del material vegetal con agentes mutagénicos. Mediante el uso de la presente invención, se proporciona a los fitomejoradores una nueva herramienta para inducir mutaciones. En consecuencia, un experto en la materia puede analizar el genoma en cuanto a fuentes de genes útiles, y en variedades que tienen características o rasgos deseados, emplear la presente invención para inducir el aumento de genes útiles, con más precisión que los agentes mutagénicos anteriores y, por tanto, acelerar y mejorar los programas de fitomejoramiento.

El polinucleótido diana de un complejo de NGA puede ser cualquier polinucleótido endógeno o exógeno para la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, un polinucleótido regulador o un ADN basura). Sin desear quedar ligado a teoría alguna, la secuencia diana debe asociarse con un PAM (motivo adyacente al protoespaciador); es decir, una secuencia corta reconocida por el sistema de NGA. Los requisitos precisos de secuencia y longitud para el PAM difieren dependiendo de la NGA utilizada, pero los PAM son normalmente secuencias de 2-5 pares de bases adyacentes al protoespaciador (es decir, la secuencia diana).

El polinucleótido diana de un sistema de NGA puede incluir varios genes y polinucleótidos asociados a una enfermedad, así como genes y polinucleótidos asociados a vías bioquímicas de señalización. Los ejemplos de polinucleótidos diana incluyen una secuencia asociada a una vía bioquímica de señalización, por ejemplo, un gen o polinucleótido asociado a la vía bioquímica de señalización. Los ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a enfermedad. Un gen o polinucleótido "asociado a enfermedad" se refiere a cualquier gen o polinucleótido que produce productos de transcripción o traducción a un nivel anómalo o en una forma anómala en células procedentes de tejidos afectados por una enfermedad en comparación tejidos o células de un control sin enfermedad. Puede ser un gen que se expresa en un nivel anormalmente alto; puede ser un gen que se expresa en un nivel anormalmente bajo, donde la expresión alterada se correlaciona con la aparición y/o la progresión de la enfermedad. Un gen asociado a enfermedad también se refiere a un gen que posee una mutación o mutaciones o variación genética que es directamente responsable o está en desequilibrio de ligamiento con un gen o genes que son responsables de la etiología de una enfermedad (por ejemplo, una mutación causal). Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar además en un nivel normal o anómalo. En algunas realizaciones, la enfermedad puede ser una enfermedad animal. En algunas realizaciones, la enfermedad puede ser una enfermedad aviar. En otras realizaciones, la enfermedad puede ser una enfermedad de mamíferos. En realizaciones adicionales, la enfermedad puede ser una enfermedad humana. Los ejemplos de genes y polinucleótidos asociados a enfermedad en seres humanos están disponibles en el McKusick-Nathans Institute of Genetic Medicine, Universidad Johns Hopkins (Baltimore, Maryland) y en el National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), disponible en Internet.

Aunque los sistemas de NGA son particularmente útiles por su relativa facilidad para dirigirse a secuencias genómicas de interés, todavía queda la cuestión de qué puede hacer la NGA para abordar una mutación causal. Un enfoque es producir una proteína de fusión entre una NGA (preferentemente una variante inactiva o de nickasa de la NGA) y una enzima editora de bases o el dominio activo de una enzima editora de bases, tal como una citidina desaminasa o un editor de bases de adenosina desaminasa (patente de EE. UU. n.º 9.840.699). En algunas realizaciones, los métodos comprenden poner en contacto una molécula de ADN con (a) una proteína de fusión que comprende una NGA en conformidad con aspectos de la invención y un polipéptido de edición de bases tal como una desaminasa; y (b) un ARNg que dirige la proteína de fusión de (a) a una secuencia de nucleótidos diana de la cadena de ADN; en donde la molécula de ADN se pone en contacto con la proteína de fusión y el ARNg en una cantidad eficaz y en condiciones adecuadas para la desaminación de una nucleobase. En algunas realizaciones, la secuencia de ADN diana comprende una secuencia asociada con una enfermedad o trastorno y en donde la desaminación de la nucleobase da como resultado una secuencia que no está asociada con una enfermedad o trastorno. En algunas realizaciones, la secuencia de ADN diana reside en un alelo de una planta de cultivo, en donde el alelo particular del rasgo de interés da como resultado una planta de menor valor agronómico. La desaminación de la nucleobase da como resultado un alelo que mejora el rasgo y aumenta el valor agronómico de la planta.

En algunas realizaciones, la secuencia de ADN comprende una mutación puntual T→C o A→G asociada con una enfermedad o trastorno, y en donde la desaminación de la base C o G mutante da como resultado una secuencia que no está asociada con una enfermedad o trastorno. En algunas realizaciones, la desaminación corrige una mutación puntual en la secuencia asociada con la enfermedad o trastorno.

En algunas realizaciones, la secuencia asociada con la enfermedad o trastorno codifica una proteína, y en donde la desaminación introduce un codón de parada en la secuencia asociada con la enfermedad o trastorno, dando como resultado un truncamiento de la proteína codificada. En algunas realizaciones, se puesta en contacto se realiza *in vivo* en un sujeto susceptible de tener, que tiene o al que se la ha diagnosticado la enfermedad o trastorno. En algunas realizaciones, la enfermedad o trastorno es una enfermedad asociada con una mutación puntual, o una mutación de base única, en el genoma. En algunas realizaciones, la enfermedad es una enfermedad genética, un cáncer, una enfermedad metabólica o una enfermedad de almacenamiento lisosómico.

IX. Composiciones farmacéuticas y métodos de tratamiento

Composiciones farmacéuticas que comprenden los polipéptidos de NGA divulgados actualmente y sus variantes activas y fragmentos, así como polinucleótidos que los codifiquen, los ARNg divulgados en el presente documento o polinucleótidos que los codifican, los sistemas divulgados en el presente documento o células que comprenden cualquiera de los polipéptidos de NGA o polinucleótidos codificantes de NGA, ARNg o polinucleótidos que codifican ARNg o los sistemas NGA, y un vehículo farmacéuticamente aceptable.

Una composición farmacéutica es una composición que se emplea para prevenir, reducir la intensidad, curar o tratar de otro modo una afección o enfermedad diana que comprende un principio activo (es decir, polipéptidos de NGA, polinucleótidos codificadores de NGA, ARNg, polinucleótidos codificadores de ARNg, sistemas de NGA o células que comprenden cualquiera de ellos) y un vehículo farmacéuticamente aceptable.

Como se usa en el presente documento, un "vehículo farmacéuticamente aceptable" se refiere a un material que no causa irritación significativa a un organismo y no deroga la actividad y las propiedades del principio activo (es decir, polipéptidos de NGA, polinucleótidos codificadores de NGA, ARNg, polinucleótidos codificadores de ARNg, sistemas de NGA o células que comprenden cualquiera de ellos). Los vehículos deben tener una pureza suficientemente elevada y una toxicidad suficientemente baja como para que puedan administrarse adecuadamente al sujeto que se va a tratar. El vehículo puede ser inerte o puede poseer beneficios farmacéuticos. En algunas realizaciones, un vehículo farmacéuticamente aceptable comprende una o más cargas, diluyentes o sustancias encapsulantes sólidos o líquidos compatibles que son adecuados para la administración a un ser humano u otro animal vertebrado. En algunas realizaciones, el vehículo farmacéuticamente aceptable no es de origen natural. En algunas realizaciones, el vehículo farmacéuticamente aceptable y el principio activo no se encuentran juntos en la naturaleza.

Las composiciones farmacéuticas utilizadas en los métodos divulgados en el presente documento pueden formularse con vehículos adecuados, excipientes y otros agentes adecuados que proporcionan una transferencia, administración, tolerancia y similares. Los expertos en la materia conocen multitud de formulaciones apropiadas. Véase, por ejemplo, Remington, The Science and Practice of Pharmacy (21ª ed. 2005). Como formulaciones adecuadas se incluyen, por ejemplo, polvos, pastas, pomadas, gelatinas, ceras, aceites, lípidos, vesículas que contienen lípidos (catiónicos o aniónicos) (tales como vesículas LIPOFECTIN, nanopartículas lipídicas, conjugados de ADN, pastas de absorción anhidras, emulsiones de aceite en agua y agua en aceite, emulsiones de Carbowax (polietilenglicoles de diversos pesos moleculares), geles semisólidos y mezclas semisólidas que contienen Carbowax. Las composiciones farmacéuticas para su uso oral o parenteral pueden prepararse en formas farmacéuticas en una dosis unitaria adecuada para ajustarse a una dosis de los principios activos. Dichas formas farmacéuticas en una dosis unitaria incluyen, por ejemplo, comprimidos, píldoras, cápsulas, inyecciones (ampollas), supositorios, etc.

En algunas realizaciones en las que las células comprenden o están modificadas con la NGA divulgada en el presente

documento, se administran al sujeto ARNg, sistemas de NGA o los polinucleótidos que los codifican, las células se administran en forma de suspensión con un vehículo farmacéuticamente aceptable. Un experto en la materia reconocerá que un vehículo farmacéuticamente aceptable para usar en una composición celular no incluirá tampones, compuestos, agentes de crioconservación, conservantes u otros agentes en cantidades que interfieran sustancialmente con la viabilidad de las células que se van a administrar al sujeto. Una formulación que comprende células puede incluir, por ejemplo, tampones osmóticos que permiten mantener la integridad de la membrana celular y, opcionalmente, nutrientes para mantener la viabilidad celular o mejorar el injerto tras la administración. Dichas formulaciones y suspensiones son conocidas por los expertos en la materia y/o pueden adaptarse para su uso con las células descritas en el presente documento usando experimentación rutinaria.

Una composición celular también se puede emulsionar o presentar como una composición de liposomas, siempre que el procedimiento de emulsificación no afecte negativamente a la viabilidad celular. Las células, y cualquier otro principio activo, pueden mezclarse con excipientes que son farmacéuticamente aceptables y compatibles con el principio activo y en cantidades adecuadas para su uso en los métodos terapéuticos descritos en el presente documento.

Los agentes adicionales incluidos en una composición celular pueden incluir sales farmacéuticamente aceptables de los componentes de la misma. Las sales farmacéuticamente aceptables incluyen las sales de adición de ácido (formadas con los grupos amino libres del polipéptido) que se forman con ácidos inorgánicos, tales como, por ejemplo, los ácidos clorhídrico o fosfórico, o ácidos orgánicos, tales como ácido acético, tartárico, mandélico y similares. Las sales formadas con los grupos carboxilo libres también proceder de bases inorgánicas, tales como, por ejemplo, sodio, potasio, amonio, calcio o férrico, y bases orgánicas tales como isopropilamina, trimetilamina, 2-etilamino-etanol, histidina, procaína y similares.

Los vehículos fisiológicamente tolerables y farmacéuticamente aceptables son bien conocidos en la técnica. Son ejemplos de vehículos líquidos las soluciones acuosas estériles que no contienen materiales además de los principios activos y agua, o contienen un tampón tal como fosfato de sodio a un valor de pH fisiológico, solución salina fisiológica o ambos, tal como solución salina tamponada con fosfato. Aún adicionalmente, los portadores acuosos pueden contener más de una sal tampón, así como sales tales como cloruro de sodio y de potasio, dextrosa, polietilenglicol y otros solutos. Las composiciones líquidas también pueden contener fases líquidas además de y excluyendo el agua. Son ejemplos de dichas fases líquidas adicionales glicerina, aceites vegetales tales como aceite de semilla de algodón y emulsiones de agua y aceite. La cantidad de un compuesto activo usado en las composiciones celulares descritas en el presente documento, que es eficaz en el tratamiento de un trastorno o afección particular, puede depender de la naturaleza del trastorno o afección, y puede determinarse mediante técnicas clínicas estándar.

Los polipéptidos de NGA divulgados en el presente documento, ARN guía, sistemas de NGA o polinucleótidos que los codifican pueden formularse con excipientes farmacéuticamente aceptables, tales como vehículos, disolventes, estabilizantes, adyuvantes, diluyentes, etc., dependiendo del modo de administración y de la forma farmacéutica particular. En algunas realizaciones, estas composiciones farmacéuticas están formuladas para alcanzar un pH fisiológicamente compatible, y oscilan entre un pH de aproximadamente 3 y un pH de aproximadamente 11, de un pH de aproximadamente 3 a un pH de aproximadamente 7, dependiendo de la formulación y vía de administración. En algunas realizaciones, el pH puede ajustarse a un intervalo de un pH de aproximadamente 5,0 a un pH de aproximadamente 8. En algunas realizaciones, las composiciones pueden comprender una cantidad terapéuticamente eficaz de al menos un compuesto descrito en el presente documento, junto con uno o más excipientes farmacéuticamente aceptables. En algunas realizaciones, las composiciones comprenden una combinación de los compuestos descritos en el presente documento, o incluyen un segundo principio activo útil en el tratamiento o la prevención del crecimiento bacteriano (por ejemplo y sin limitación, agentes antibacterianos o antimicrobianos), o incluyen una combinación de reactivos de la presente divulgación.

Los excipientes adecuados incluyen, por ejemplo, moléculas vehículo que incluyen macromoléculas grandes, de metabolización lenta, tales como proteínas, polisacáridos, ácidos polilácticos, ácidos poliglicólicos, aminoácidos poliméricos, copolímeros aminoácídicos y partículas víricas inactivas. Otros excipientes ilustrativos pueden incluir antioxidantes (por ejemplo y sin limitación, ácido ascórbico), agentes quelantes (por ejemplo y sin limitación, EDTA), hidratos de carbono (por ejemplo y sin limitación, dextrina, hidroxialquilcelulosa e hidroxialquilmethylcelulosa), ácido esteárico, líquidos (por ejemplo y sin limitación, aceites, agua, solución salina, glicerol y etanol), agentes humectantes o emulsionantes, sustancias tamponantes del pH y similares.

En algunas realizaciones, las formulaciones se presentan en envases de dosis individual o de multidosis, por ejemplo ampollas y viales cerrados herméticamente y se pueden almacenar en un estado criodesecado (liofilizado) que requiere la adición del vehículo líquido estéril, por ejemplo, solución salina, agua para inyección, una espuma semilíquida, o gel, inmediatamente antes de su uso. Pueden prepararse soluciones y suspensiones para inyección extemporánea a partir de polvos, gránulos y comprimidos estériles del tipo descrito anteriormente. En algunas realizaciones, el principio activo se disuelve en una solución líquida tamponada que se congela en un envase monodosis o multidosis y posteriormente se descongela para su inyección o se mantiene/estabiliza en refrigeración hasta su uso.

El agente o agentes terapéuticos pueden estar contenidos en sistemas de liberación controlada. Con el fin de prolongar

el efecto de un fármaco, a menudo es deseable ralentizar la absorción del fármaco de la inyección subcutánea, intratecal o intramuscular. Esto puede lograrse mediante el uso de una suspensión líquida de material cristalino o amorfo con escasa solubilidad en agua. La velocidad de absorción del fármaco depende, por tanto, de su velocidad de disolución que, a su vez, puede depender del tamaño del cristal y de la forma cristalina. Como alternativa, la absorción retardada de una forma de fármaco administrada por vía parenteral se logra disolviendo o suspendiendo el fármaco en un vehículo oleoso. En algunas realizaciones, el uso de un implante de liberación sostenida a largo plazo puede ser especialmente adecuado para el tratamiento de afecciones crónicas. Los implantes de liberación sostenida a largo plazo son bien conocidos por los expertos en la materia.

Se proporcionan métodos para tratar una enfermedad en un sujeto que lo necesita. Los métodos comprenden la administración a un sujeto que lo necesite de una cantidad eficaz de un polipéptido de NGA divulgado en el presente documento, una variante activa o un fragmento del mismo o de un polinucleótido que lo codifique, un ARNg divulgado en el presente documento o un polinucleótido que lo codifica, un sistema de NGA divulgado en el presente documento o una célula modificada por o que comprende cualquiera de estas composiciones.

En algunas realizaciones, el tratamiento comprende *edición génica in vivo* mediante la administración de un polipéptido NGA divulgado en el presente documento, ARNg o sistema de NGA o polinucleótido(s) que lo codifique(n), en donde el tratamiento no es un proceso para modificar la identidad genética de la línea germinal de un ser humano. En algunas realizaciones, el tratamiento comprende la edición génica *ex vivo* en la que las células se modifican genéticamente *ex vivo* con un polipéptido NGA divulgado en el presente documento, ARNg o sistema de NGA o polinucleótido(s) que lo codifica(n) y, a continuación, las células modificadas se administran a un sujeto. En algunas realizaciones, las células modificadas genéticamente proceden del sujeto al que se le administran las células modificadas, y las células trasplantadas se denominan aquí autólogas. En algunas realizaciones, las células modificadas genéticamente proceden de un sujeto diferente (es decir, donante) dentro de la misma especie que el sujeto al que se administran las células modificadas (es decir, receptor), y las células trasplantadas se denominan aquí alogénicas. En algunos ejemplos descritos en el presente documento, las células pueden ampliarse en cultivo antes de su administración a un sujeto que lo necesite.

En algunas realizaciones, la enfermedad a tratar con las composiciones divulgadas en el presente documento es una que puede tratarse con inmunoterapia, tal como con un linfocito T con receptor de antígeno quimérico (CAR). Dichas enfermedades incluyen, pero sin limitación, cáncer. En algunas realizaciones, la enfermedad que se va a tratar con las composiciones divulgadas en el presente documento está asociada a una mutación causal. Como se usa en el presente documento, una "mutación causal" se refiere a un nucleótido concreto, nucleótidos o secuencia de nucleótidos en el genoma que contribuye a la gravedad o presencia de una enfermedad o trastorno en un sujeto. La corrección de la mutación causal conduce a la mejora de al menos un síntoma resultante de una enfermedad o trastorno. En algunas realizaciones, la mutación causal es adyacente a un sitio PAM reconocido por un NGA divulgada en el presente documento. La mutación causal puede corregirse con una NGA divulgada en el presente documento o con un polipéptido de fusión que comprende una NGA divulgada en el presente documento y un polipéptido editor de bases (es decir, un editor de bases). Ejemplos no limitantes de enfermedades asociadas a una mutación causal incluyen fibrosis quística, síndrome de Hurler, ataxia de Friedreich, enfermedad de Huntington y anemia falciforme. En algunas realizaciones, la enfermedad a tratar con las NGA divulgadas en el presente documento es una enfermedad de las enumeradas en la Tabla 11. El Instituto de Medicina Genética McKusick-Nathans ofrece otros ejemplos no limitativos de genes y mutaciones asociados a enfermedades, Universidad Johns Hopkins (Baltimore, Maryland) y en el National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), disponible en Internet.

Como se usa en el presente documento, "tratamiento" o "tratar", o "paliar" o "mejorar" se usan indistintamente. Estos términos se refieren a un enfoque para obtener resultados beneficiosos o deseados que incluyen, aunque no de forma limitativa, un beneficio terapéutico y/o un beneficio profiláctico. Por beneficio terapéutico se entiende cualquier mejora o efecto terapéuticamente relevante en una o más enfermedades, afecciones o síntomas en tratamiento. Para el beneficio profiláctico, las composiciones pueden administrarse a un sujeto con riesgo de desarrollar una enfermedad particular, afección o síntoma, o a un sujeto que informa uno o más de los síntomas fisiológicos de una enfermedad, aunque la enfermedad, la afección o el síntoma puede no haberse manifestado todavía.

La expresión "cantidad eficaz" o "cantidad terapéuticamente eficaz" se refiere a la cantidad de un agente que es suficiente para conseguir resultados beneficiosos o deseados. La cantidad terapéuticamente eficaz puede variar dependiendo de uno o más de: el sujeto y la patología que se esté tratando, el peso y la edad del sujeto, la gravedad de la patología, la forma de administración y similares, lo que puede determinar fácilmente el experto en la materia. La dosis específica puede variar dependiendo de uno o más de: el agente particular elegido, la pauta posológica a seguir, si el compuesto se administra en combinación con otros compuestos, el momento de la administración y el sistema de administración en el que se realiza.

El término "administrar" se refiere a la colocación de un principio activo en un sujeto, mediante un método o una vía que de como resultado la localización, al menos parcial, del principio activo introducido en un lugar deseado, tal como un lugar de lesión o reparación, de manera que se produzcan los efectos deseados. En las realizaciones en donde se administran células, las células pueden administrarse mediante cualquier vía apropiada que suponga una administración a una ubicación deseada en el sujeto donde al menos una parte de las células o componentes de las

células implantadas permanecen viables. El período de viabilidad de las células después de su administración a un sujeto, puede ser tan corto como de algunas horas, por ejemplo, veinticuatro horas, a algunos días, hasta varios años, o incluso toda la vida del paciente, es decir, injerto a largo plazo. Por ejemplo, en algunos aspectos descritos en el presente documento, una cantidad eficaz de células fotorreceptoras o células progenitoras de la retina se administra por vía sistémica, tal como una vía intraperitoneal o intravenosa.

En algunas realizaciones, la administración comprende la administración por vía vírica. En algunas realizaciones, la administración comprende la administración por electroporación. En algunas realizaciones, la administración comprende la administración por vía de nanopartículas. En algunas realizaciones, la administración comprende la administración por vía de liposomas. Puede utilizarse cualquier vía de administración eficaz para administrar una cantidad eficaz de una composición farmacéutica descrita en el presente documento. En algunas realizaciones, la administración comprende la administración mediante un método seleccionado del grupo que consiste en: intravenosa, subcutánea, intramuscular, vía oral, vía rectal, por aerosol, parenteral, oftálmica, pulmonar, transdérmica, por vía vaginal, ótica, por vía nasal y por administración tópica, o cualquier combinación de las mismas. En algunas realizaciones, para la administración de células, se utiliza la administración por inyección o infusión.

Como se usa en el presente documento, el término "sujeto" se refiere a cualquier individuo para el que se desee un diagnóstico, tratamiento o terapia. En algunas realizaciones, el sujeto es un animal. En algunas realizaciones, el sujeto es un mamífero. En algunas realizaciones, el sujeto es un ser humano.

La eficacia de un tratamiento puede ser determinada por el clínico experto. Sin embargo, un tratamiento se considera un "tratamiento eficaz", si alguno o todos los signos o síntomas de una enfermedad o trastorno se alteran de forma beneficiosa (por ejemplo, disminuyen al menos un 10 %) o mejoran o alivian otros síntomas o marcadores de enfermedad clínicamente aceptados. La eficacia también puede medirse mediante la incapacidad de un individuo para empeorar según la evaluación de la hospitalización o la necesidad de intervenciones médicas (p. ej., la progresión de la enfermedad se detiene o al menos se ralentiza). Los expertos en la materia conocen métodos para medir estos indicadores. El tratamiento incluye: (1) inhibir la enfermedad, por ejemplo, detener o ralentizar la progresión de los síntomas; o (2) aliviar la enfermedad, por ejemplo, provocando regresión de los síntomas; y (3) prevenir o reducir la probabilidad de desarrollo de los síntomas.

A. Modificación de mutaciones causales mediante la edición de bases

Un ejemplo de una enfermedad de herencia genética que podría corregirse utilizando un enfoque que se basa en una proteína de fusión de NGA-editor de bases de la invención es el síndrome de Hurler. El síndrome de Hurler, también conocido como MPS-1, es el resultado de una deficiencia de α -L-iduronidasa (IDUA) que es el resultado de una enfermedad de almacenamiento lisosómico caracterizada a nivel molecular por la acumulación de dermatán sulfato y heparán sulfato en los lisosomas. Esta enfermedad es generalmente un trastorno genético hereditario provocado por mutaciones en el gen IDUA que codifica la α -L-iduronidasa. Las mutaciones de IDUA comunes son W402X y Q70X, mutaciones terminadoras que dan como resultado la terminación prematura de la traducción. Dichas mutaciones se abordan bien mediante enfoques precisos de edición genómica (PEG), dado que la inversión de un solo nucleótido, por ejemplo mediante un enfoque de edición de bases, restablecería la secuencia codificante de tipo silvestre y daría como resultado la expresión de proteínas controlada por los mecanismos reguladores endógenos del locus genético. Además, dado que se sabe que los heterocigóticos son asintomáticos, una terapia con PGE dirigida a una de estas mutaciones sería útil para una gran proporción de pacientes con esta enfermedad, ya que sólo es necesario corregir uno de los alelos mutados (Bunge *et al.* (1994) Hum. Mol. Genet. 3(6): 861-866).

Los tratamientos actuales para el síndrome de Hurler incluyen terapia de reemplazo enzimático y trasplantes de médula ósea (Vellodi *et al.* (1997) Arq. Dis. Child. 76(2): 92-99; Peters *et al.* (1998) Blood 91(7): 2601-2608). Si bien la terapia de reemplazo enzimático ha tenido un efecto espectacular en la supervivencia y la calidad de vida de los pacientes con síndrome de Hurler, este enfoque requiere infusiones semanales costosas y que requieren mucho tiempo. Los enfoques adicionales incluyen la administración del gen IDUA en un vector de expresión o la inserción del gen en un locus altamente expresado tal como el de la seroalbúmina (patente de EE. UU. n.º 9.956.247). Sin embargo, estos enfoques no restablecen el locus IDUA original a la secuencia codificante correcta. Una estrategia de edición genómica tendría una serie de ventajas, la más notable que la regulación de la expresión génica estaría controlada por los mecanismos naturales presentes en individuos sanos. Además, el uso de la edición de bases no requiere provocar roturas de ADN bicatenarias, lo que podría conducir a grandes reordenamientos cromosómicos, a muerte celular u oncogenicidad por la alteración de los mecanismos de supresión tumoral. Una estrategia general puede dirigirse hacia el uso de proteínas de fusión de NGA-editor de base de la invención para apuntar y corregir ciertas mutaciones causantes de enfermedades en el genoma humano. Se apreciará que también pueden aplicarse enfoques similares para tratar enfermedades que pueden corregirse mediante la edición de bases. Se apreciará además que enfoques similares para dirigirse a mutaciones causantes de enfermedades en otras especies, especialmente animales domésticos o de cría comunes, también pueden desplegarse utilizando las NGA de la invención. Las mascotas y el ganado domésticos más comunes incluyen perros, gatos, caballos, cerdos, vacas, ovejas, pollos, burros, serpientes, hurones y peces, incluidos el salmón y las gambas.

B. Modificación de mutaciones causales mediante delección dirigida

Las NGA de la invención también podrían ser útiles en enfoques terapéuticos en seres humanos donde la mutación causal sea más complicada. Por ejemplo, algunas enfermedades tales como la ataxia de Friedreich y la enfermedad de Huntington son el resultado de un aumento significativo en las repeticiones de un motivo de tres nucleótidos en una región particular de un gen, lo que afecta a la capacidad de la proteína expresada para funcionar o expresarse. La ataxia de Friedreich (FRDA, del inglés *Friedreich's Ataxia*) es una enfermedad autosómica recesiva que provoca una degeneración progresiva del tejido nervioso en la médula espinal. Los niveles reducidos de la proteína frataxina (FXN) en las mitocondrias provocan daños oxidativos y deficiencias de hierro a nivel celular. La expresión reducida de FXN se ha vinculado a una expansión del triplete GAA dentro del intrón 1 del gen FXN somático y de la línea germinal. En pacientes con FRDA, la repetición GAA frecuentemente consiste en más de 70, a veces incluso más de 1000 (más comúnmente 600-900) tripletes, mientras que los individuos no afectados tienen aproximadamente 40 repeticiones o menos (Pandolfo *et al.* (2012) *Handbook of Clinical Neurology* 103: 275-294; Campuzano *et al.* (1996) *Science* 271: 1423-1427; Pandolfo (2002) *Adv. Exp. Med. Biol.* 516: 99-118).

La expansión de la secuencia de repetición trinucleotídica que provoca la ataxia de Friedreich (FRDA) se produce en un locus genético definido dentro del gen FXN, conocida como la región de inestabilidad de FRDA. Se pueden usar nucleasas guiadas por ARN (las NGA) para escindir la región de inestabilidad en células de pacientes con FRDA. Este enfoque precisa 1) una NGA y una secuencia de ARN de guía que pueda programarse para dirigirse al alelo en el genoma humano; y 2) un enfoque de administración para la NGA y la secuencia de guía. Muchas nucleasas utilizadas para la edición genómica, tales como la comúnmente utilizada nucleasa Cas9 de *S. pyogenes* (SpCas9), son demasiado grandes para empaquetarse en vectores de virus adenoasociado (AAV), especialmente cuando se considera la longitud del gen de la SpCas9 y el ARN de guía además de otros elementos genéticos necesarios para casetes de expresión funcionales. Esto hace que el enfoque que utiliza SpCas9 sea más difícil.

Ciertas nucleasas guiadas por ARN de la invención son muy adecuadas para su empaquetamiento en un vector AAV junto con un ARN guía. El empaquetamiento de dos ARN de guía posiblemente podría precisar un segundo vector, pero este enfoque aún se compara favorablemente con lo que se precisaría de una nucleasa más grande tal como SpCas9, que puede precisar dividir la secuencia proteica entre dos vectores. La presente invención abarca una estrategia que utiliza NGA de la invención en la que se elimina una región de inestabilidad genómica. Dicha estrategia es aplicable a otras enfermedades y trastornos que tienen una base genética similar, tales como la enfermedad de Huntington. Estrategias similares utilizando NGA de la invención también pueden ser aplicables a enfermedades y trastornos similares en animales no humanos de importancia agronómica o económica, incluyendo perros, gatos, caballos, cerdos, vacas, ovejas, pollos, burros, serpientes, hurones y peces, incluidos el salmón y las gambas.

C. Modificación de mutaciones causales mediante mutagénesis dirigida

Las NGA de la invención también podrían ser para introducir mutaciones disruptivas que pueden dar lugar a un efecto beneficioso. Los defectos genéticos en los genes que codifican la hemoglobina, particularmente la cadena de beta globina (el gen HBB), pueden ser responsable de una serie de enfermedades conocidas como hemoglobinopatías, incluyendo la anemia falciforme y las talasemias.

En seres humanos adultos, la hemoglobina es un heterotetrámero que comprende dos cadenas de globina de tipo alfa (α) y dos cadenas de globina de tipo beta (β), y 4 grupos hemo. En los adultos, el tetrámero $\alpha_2\beta_2$ se denomina hemoglobina A (HbA) o hemoglobina del adulto. Típicamente, las cadenas de globina alfa y beta se sintetizan en una relación aproximada de 1:1 y esta relación parece ser crítica en términos de estabilización de la hemoglobina y los glóbulos rojos (GR). En un feto en desarrollo, se produce una forma diferente de hemoglobina, la hemoglobina fetal (HbF), que tiene una mayor afinidad de unión por el oxígeno que la hemoglobina A, de modo que el oxígeno puede suministrarse al sistema del bebé a través del torrente sanguíneo de la madre. La hemoglobina fetal también contiene dos cadenas de globina α , pero en lugar de las cadenas de globina β del adulto, tiene dos cadenas de globina gamma (γ) fetal (es decir, la hemoglobina fetal es $\alpha_2\gamma_2$). La regulación del cambio de la producción de globina gamma a beta es bastante compleja e implica principalmente una regulación a la baja de la transcripción de la globina gamma con una regulación al alza simultánea de la transcripción de globina beta. Aproximadamente a las 30 semanas de gestación, la síntesis de globina gamma en el feto comienza a disminuir mientras que aumenta la producción de globina beta. Aproximadamente a los 10 meses de edad, la hemoglobina del recién nacido es casi toda $\alpha_2\beta_2$, aunque algo de HbF persiste hasta la edad adulta (aproximadamente entre el 1-3 % de la hemoglobina total). En la mayoría de los pacientes con hemoglobinopatías, los genes que codifican la globina gamma permanecen presentes, pero la expresión es relativamente baja debido a la represión génica normal que se produce en torno al parto, como se describió anteriormente.

La anemia drepanocítica está provocada por una mutación V6E en el gen de la globina β (HBB) (GAG a GTG a nivel del ADN), donde la hemoglobina resultante se denomina "hemoglobina S" o "HbS". En condiciones de menor oxígeno, las moléculas de HbS se agregan y forman precipitados fibrosos. Estos agregados provocan la anomalía o "falciformación" de los GR, dando como resultado una pérdida de flexibilidad de las células. Los GR con falciformación ya no pueden introducirse en los lechos capilares y pueden dar como resultado una crisis vasooclusiva en pacientes con drepanocitos. Además, los glóbulos rojos con falciformación son más frágiles que los GR normales y tienden a la hemólisis, provocando finalmente anemia en el paciente.

El tratamiento y tratamiento de los pacientes con anemia drepanocítica es una propuesta de por vida que implica tratamiento con antibióticos, tratamiento del dolor y transfusiones durante los episodios agudos. Un enfoque es el uso de hidroxiurea, que ejerce sus efectos en parte aumentando la producción de globina gamma. Aún se desconocen los efectos secundarios a largo plazo del tratamiento crónico con hidroxiurea, sin embargo, el tratamiento produce efectos secundarios no deseados y puede tener una eficacia variable de un paciente a otro. A pesar de un aumento en la eficacia de los tratamientos de los drepanocitos, la esperanza de vida de los pacientes todavía se sitúa entre mediados y finales de los 50 años y las morbilidades asociadas a la enfermedad tienen un profundo impacto en la calidad de vida del paciente.

Las talasemias (talasemias alfa y talasemia beta) también son enfermedades relacionadas con la hemoglobina y normalmente implican una expresión reducida de las cadenas de globina. Esto puede tener lugar a través de mutaciones en las regiones reguladoras de los genes o de una mutación en una secuencia codificante de globina que dé como resultado una expresión reducida o niveles reducidos de proteína globina funcional. El tratamiento de las talasemias implica habitualmente transfusiones de sangre y terapia de quelación del hierro. Los trasplantes de médula ósea también se utilizan para el tratamiento de personas con talasemias graves si se puede identificar un donante adecuado, pero este procedimiento puede tener riesgos importantes.

Un enfoque que se ha propuesto para el tratamiento tanto de la anemia falciforme (AF) como de las talasemias beta es aumentar la expresión de globina gamma para que la HbF sustituya funcionalmente a la hemoglobina del adulto aberrante. Como se ha mencionado anteriormente, se cree que el tratamiento de pacientes con AD con hidroxiurea es satisfactorio en parte debido a su efecto sobre el aumento de la expresión de globina gamma (DeSimone (1982) Proc Natl Acad Sci USA 79(14):4428-31; Ley, *et al.*, (1982) N. Engl. J. Medicine, 307: 1469-1475; Ley, *et al.*, (1983) Blood 62: 370-380; Constantoulakis *et al.*, (1988) Blood 72(6):1961-1967). Aumentar la expresión de HbF implica la identificación de genes cuyos productos desempeñan un papel en la regulación de la expresión de globina gamma. Uno de esos genes es BCL11A. BCL11A codifica una proteína de dedos de Cinc que se expresa en células precursoras eritroides adultas y la regulación por disminución de su expresión conduce a un aumento en la expresión de globina gamma (Sankaran *et al* (2008) Science 322: 1839, incorporada en el presente documento por referencia). Se ha propuesto el uso de un ARN inhibidor dirigido al gen BCL11A (por ejemplo, publicación de patente de EE. UU. 2011/0182867), pero esta tecnología tiene varios posibles inconvenientes, incluido que es posible que no se logre una atenuación génica completa, que la administración de dichos ARN puede ser problemático y que los ARN deben estar presentes continuamente, precisándose múltiples tratamientos de por vida.

Las NGA de la invención pueden utilizarse para dirigirse a la región potenciadora de BCL11A para interrumpir la expresión de BCL11A, aumentando así la expresión de la globina gamma. Esta alteración específica se puede lograr mediante la unión de extremos no homólogos (NHEJ), mediante lo cual una NGA de acuerdo con la invención se dirige a una secuencia particular dentro de la región potenciadora de BCL11A, produce una rotura bicatenaria y la maquinaria de la célula repara la rotura, por lo general, introduciendo simultáneamente mutaciones nocivas. De manera similar a lo que se describe para otras dianas de enfermedades, Las NGA de la invención pueden tener ventajas sobre otras NGA conocidas debido a su tamaño relativamente pequeño, lo que facilita el empaquetamiento de los casetes de expresión para la NGA y su ARN de guía en un único vector de AAV para la administración *in vivo*. Estrategias similares utilizando NGA de la invención también pueden ser aplicables a enfermedades y trastornos similares tanto en seres humanos como en animales no humanos de importancia agronómica o económica.

X. Células que comprenden una modificación genética de polinucleótidos

En el presente documento se proporcionan células y organismos no humanos que comprenden una secuencia diana de interés que se ha modificado usando un procedimiento mediado por una NGA, ARNcr y/o ARNtrat como se describe en el presente documento, en donde el proceso no es un proceso para modificar la identidad genética de la línea germinal de seres humanos. La NGA comprende la secuencia de aminoácidos de SEQ ID NO: 83, o una variante activa o fragmento de la misma de acuerdo con aspectos de la invención. En diferentes realizaciones, el ARN de guía comprende una secuencia de repetición de CRISPR que comprende la secuencia de nucleótidos de las SEQ ID NO: 2, 9, 16, 23, 30, 37, 44, 51, 57, 64, 71, 77, 84, 90, 97, 104, 111, 118, o 124, o una variante activa o fragmento del mismo. En realizaciones particulares, el ARN de guía comprende un ARNtrat que comprende la secuencia de nucleótidos de las SEQ ID NO: 3, 10, 17, 24, 31, 38, 45, 52, 58, 65, 72, 78, 85, 91, 98, 105, 112, 119, o 125, o una variante activa o fragmento del mismo. El ARN de guía del sistema puede ser un ARN de guía único o un ARN de guía doble.

Las células modificadas pueden ser eucariotas (por ejemplo, células de mamífero, vegetales o de insecto) o procariotas. También se proporcionan órganos y embriones no humanos que comprenden al menos una secuencia de nucleótidos que ha sido modificada mediante un procedimiento que utiliza una NGA, un ARNcr y/o un ARNtrat como se describe en el presente documento. Las células, organismos, órganos y embriones no humanos modificados genéticamente pueden ser heterocigotos u homocigotos para la secuencia de nucleótidos modificada.

La modificación cromosómica de la célula, organismo, órgano o embrión no humano puede provocar una expresión alterada (regulación al alza o a la baja), inactivación, o la expresión de un producto proteico alterado o una secuencia

integrada. En las realizaciones en donde la modificación cromosómica da como resultado la inactivación de un gen o la expresión de un producto proteico no funcional, la célula, organismo, o embrión no humano modificada genéticamente se denominan "atenuados". El fenotipo de atenuación génica puede ser el resultado de una mutación por delección (es decir, delección de al menos un nucleótido), una mutación de inserción (es decir, la inserción de al menos un nucleótido), o una mutación sin sentido (es decir, la sustitución de al menos un nucleótido de forma que se introduce un codón de terminación).

Como alternativa, la modificación cromosómica de una célula, organismo, orgánulo o embrión no humano puede producir una "inserción génica", que da como resultado la integración cromosómica de una secuencia de nucleótidos que codifica una proteína. En algunos de estos ejemplos, la secuencia codificante se integra en el cromosoma de manera que la secuencia cromosómica que codifica la proteína de tipo silvestre está inactivada, pero se expresa la proteína introducida exógenamente.

En otros ejemplos, la modificación cromosómica da como resultado la producción de un producto proteico variante. El producto proteico variante expresado puede tener al menos una sustitución de aminoácido y/o la adición o delección de al menos un aminoácido. El producto proteico variante codificado por la secuencia cromosómica alterada puede presentar características o actividades modificadas en comparación con la proteína de tipo silvestre, que incluyen, pero sin limitación, una actividad enzimática o especificidad de sustrato alterada.

En otros ejemplos más, la modificación cromosómica puede dar como resultado un patrón de expresión alterado de una proteína. Como ejemplo no limitante, las alteraciones cromosómicas en las regiones reguladoras que controlan la expresión de un producto proteico pueden dar como resultado la sobreexpresión o la regulación a la baja del producto proteico o un patrón alterado de la expresión temporal o tisular.

Las células modificadas pueden convertirse en un organismo, tal como una planta, en conformidad con las formas convencionales. Véase, por ejemplo, McCormick *et al.* (1986) Plant Cell Reports 5:81-84. A continuación, estas plantas pueden cultivarse y polinizarse con la misma cepa modificada o con cepas diferentes, y el híbrido resultante tener la modificación genética. La presente invención permite obtener semillas modificadas genéticamente. También se facilitan la progenie, variantes y mutantes de las plantas regeneradas, siempre que estas partes comprendan la modificación genética. Además se proporciona un producto o subproducto vegetal procesado que conserva la modificación genética, incluyendo, por ejemplo, harina de soja.

Los métodos proporcionados en el presente documento pueden usarse para la modificación de cualquier especie vegetal, incluyendo, aunque no de forma limitativa, monocotiledóneas o dicotiledóneas. Los ejemplos de plantas de interés incluyen, pero sin limitación, gramíneas (maíz), sorgo, trigo, girasol, tomate, crucíferas, pimientos, patata, algodón, arroz, soja, remolacha azucarera, caña de azúcar, tabaco, cebada y colza, Brassica sp., alfalfa, centeno, mijo, cártamo, cacahuètes, batata, mandioca, café, coco, piña, árboles de cítricos, cacao, té, plátano, aguacate, higuera, guayaba, mango, aceituna, papaya, anacardo, macadamia, almendra, avena, verduras, plantas ornamentales y coníferas.

Las verduras incluyen, pero sin limitación, tomates, lechuga, judías verdes, habas de Lima, guisantes y miembros del género *Curcumis* tales como pepino, melón cantalupo y melón. Las plantas ornamentales incluyen, pero sin limitación, azalea, hortensia, hibisco, rosas, tulipán, narcisos, petunias, clavel, flor de pascua y crisantemo. Preferentemente, las plantas de la presente invención son plantas de cultivo (por ejemplo, maíz, sorgo, trigo, girasol, tomate, crucíferas, pimientos, patata, algodón, arroz, soja, remolacha azucarera, caña de azúcar, tabaco, cebada, colza, etc.).

Los métodos proporcionados en el presente documento también pueden utilizarse para modificar genéticamente cualquier especie procariota, incluyendo, pero sin limitación, arqueas y bacterias (por ejemplo, *Bacillus* sp., *Klebsiella* sp., *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., *Agrobacterium*, *Lactobacillus* sp.).

Los métodos proporcionados en el presente documento pueden utilizarse para modificar genéticamente cualquier especie eucariota o sus células, que incluyen, aunque no de forma limitativa, animales (por ejemplo, mamíferos, insectos, peces, aves y reptiles), hongos, amebas, algas y levaduras. En algunas realizaciones, la célula que se modifica mediante los métodos divulgados en el presente documento incluye células de origen hematopoyético, tales como células del sistema inmunitario, incluidas, entre otras, linfocitos B, linfocitos T, linfocitos citolíticos naturales (NK), células madre pluripotenciales, células madre pluripotenciales inducidas, linfocitos T receptores de antígeno quimérico (CAR-T), monocitos, macrófagos y células dendríticas.

Las células que se han modificado pueden introducirse en un organismo. Estas células podrían proceder del mismo organismo (por ejemplo, una persona) en el caso de los trasplantes celulares autólogos, en donde las células se modifican en un enfoque *ex vivo*. Como alternativa, las células proceden de otro organismo de la misma especie (por ejemplo, otra persona) en el caso de los trasplantes celulares alogénicos.

Los artículos "un" y "una" se usan en el presente documento para hacer referencia a uno o más de uno (es decir, al menos uno) del objeto gramatical del artículo. A modo de ejemplo, "un polipéptido" significa uno o más polipéptidos.

Todas las publicaciones y solicitudes de patente mencionadas en la memoria descriptiva son indicativas del nivel de los expertos en la materia a la que pertenece la presente divulgación.

- 5 Aunque la invención anterior se ha descrito con cierto nivel de detalle a modo de ilustración y ejemplo con fines de claridad de comprensión, será obvio que pueden practicarse ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

Los siguientes ejemplos se ofrecen a modo de ilustración y no a modo de limitación.

10

Ejemplos

Ejemplo 1. Identificación de nucleasas guiadas por ARN

- 15 Se identificaron diecinueve nucleasas guiadas por ARN (NGA) asociadas a CRISPR distintas y se describen a continuación en la Tabla 1. La Tabla 1 proporciona el nombre de cada NGA, su secuencia de aminoácidos, la fuente de la que procede y las secuencias procesadas de ARNcr y ARNtrat (véanse los métodos de identificación en el Ejemplo 2). La Tabla 1 proporciona además una secuencia de ARN guía genérico único (ARNgu), donde la poli-N indica la ubicación de la secuencia espaciadora que determina la secuencia diana de ácido nucleico del ARNgu. Para los sistemas de NGA APG06622, APG02787 y APG06248, la secuencia conservada en la base del tallo de la horquilla del ARNtrat es UNANNA (SEQ ID NO: 129). Para APG06007, APG09344 y APG07991, la secuencia en la misma ubicación es UNANNU (SEQ ID NO: 130). Para APG02874, APG03850 y APG07553, la secuencia en la misma ubicación es UNANNG (SEQ ID NO: 131). Para los sistemas de NGA APG03031, APG09208, APG05586, APG08770, APG03021, APG06015, APG01868 y APG02998, la secuencia conservada en la base del tallo de la horquilla del ARNtrat es UNANNC (SEQ ID NO: 132). Para APG08167 y APG01604, la secuencia en la misma ubicación es CNANNC (SEQ ID NO: 133).
- 20
- 25

Tabla 1: Sumario de las SEQ ID y de los sistemas asociados a CRISPR

ID de NGA	SEQ ID NO.	Fuente	secuencia de repetición de ARNcr (SEQ ID NO:)	ARNtrat (SEQ ID NO:)	estructura del ARNgu (SEQ ID NO)
APG06622	1	<i>Pedobacter</i> sp.	2	3	4
APG02787	8	<i>Chitinophaga</i> sp.	9	10	11
APG06248	15	<i>Mucilaginibacter</i> sp.	16	17	18
APG06007	22	<i>Acidovorax</i> sp.	23	24	25
APG02874	29	<i>Bacillus</i> sp.	30	31	32
APG03850	36	<i>Bacillus</i> sp.	37	38	39
APG07553	43	<i>Bacillus</i> sp.	44	45	46
APG03031	50	<i>Chryseobacterium</i> sp.	51	52	53
APG09208	56	<i>Bacillus</i> sp.	57	58	59
APG05586	63	<i>Enterococcus</i> sp.	64	65	66
APG08770	70	<i>Enterococcus</i> sp.	71	72	73
APG08167	76	<i>Staphylococcus</i> sp.	77	78	79
APG01604	83	<i>Staphylococcus</i> sp.	84	85	86
APG03021	89	<i>Streptococcus</i> sp.	90	91	92
APG06015	96	<i>Pediococcus</i> sp.	97	98	99
APG09344	103	<i>Weissella</i> sp.	104	105	106
APG07991	110	<i>Enterococcus</i> sp.	111	112	113
APG01868	117	<i>Enterococcus</i> sp.	118	119	120
APG02998	123	<i>Enterococcus</i> sp.	124	125	126

- 30 Ejemplo 2: Identificación de ARN de guía y construcción de ARNgu

Los cultivos de bacterias que expresan de forma nativa el sistema de nucleasa guiado por ARN bajo investigación se cultivaron hasta la fase semilogarítmica (DO600 de ~0,600), se sedimentaron y se congelaron instantáneamente. El ARN se aisló de los sedimentos utilizando un kit mirVANA miRNA Isolation (Life Technologies, Carlsbad, CA) y se prepararon bibliotecas de secuenciación a partir del ARN aislado utilizando un kit NEBNext Small RNA Library Prep (NEB, Beverly, MA). La preparación de la biblioteca se fraccionó en un gel de poliacrilamida al 6 % para capturar las especies de ARN inferiores a 200 nt y detectar los ARNcr y los ARNtrat, respectivamente. Un proveedor de servicios (MoGene, St. Louis, MO) realizó una secuenciación profunda (75 pares de bases del extremo) en un Next Seq 500 (kit de alto rendimiento). La calidad de las lecturas se recortó con Cutadapt y se mapearon en genomas de referencia utilizando Bowtie2. Se escribió una cartera de RNAseq personalizada en Python para detectar los transcritos de ARNcr y ARNtrat. Los límites del ARNcr procesado se determinaron mediante la cobertura de secuencia de la matriz espaciadora de repetición nativa. La porción antirrepetición del ARNtrat se identificó utilizando parámetros de BLASTn permisivos. La profundidad de la secuenciación de ARN confirmó los límites del ARNtrat procesado mediante la identificación del transcrito que contenía la antirrepetición. La curación manual de los ARN se realizó mediante la predicción de la estructura secundaria mediante NUPACK, un programa informático de plegado de ARN. Los casetes de ARNgu se prepararon mediante síntesis de ADN y, por lo general, se diseñaron de la siguiente manera (5'→3'): Secuencia espaciadora de 20-30 pb, unida operativamente en su extremo 3' a la porción de repetición procesada del ARNcr, unida de forma operable a un enlazador no complementario de 4 pb (AAAG; SEQ ID NO: 249) unido operativamente en su extremo 3' al ARNtrat procesado. También pueden utilizarse otros enlazadores no complementarios de 4 pb.

Para los ensayos *in vitro*, los ARNgu se sintetizaron mediante transcripción *in vitro* de los casetes de ARNgu con un kit GeneArt™ Precision gRNA Synthesis Kit (ThermoFisher). Las secuencias de ARNcr y ARNtrat procesados para cada uno de los polipéptidos de NGA se identifican y se exponen en la Tabla 1. Consúltense a continuación los ARNgu contruidos para las bibliotecas de PAM 1 y 2.

Ejemplo 3: Determinación de los requisitos de PAM para cada NGA

Los requisitos de PAM para cada NGA se determinaron utilizando un ensayo de empobrecimiento de PAM esencialmente adaptado a partir de Kleinstiver *et al.* (2015) Nature 523:481-485 y Zetsche *et al.* (2015) Cell 163:759-771. Brevemente, se generaron dos bibliotecas de plásmidos (L1 y L2) en una cadena principal de pUC18 (ampR), cada una de las cuales contenía una secuencia protoespaciadora (diana) de 30 pb distinta flanqueada por 8 nucleótidos aleatorios (es decir, la región PAM). La secuencia diana y la región PAM flanqueante de la biblioteca 1 y la biblioteca 2 para cada NGA se exponen en la Tabla 2.

Las bibliotecas se electroporaron por separado en células de *E. coli* BL21 (DE3) que alojaban vectores de expresión pRSF-1b que contenían una NGA de la invención (codones optimizados para *E. coli*) junto con un ARNgu afín que contenía una secuencia espaciadora correspondiente al protoespaciador en L1 o L2. Se utilizó suficiente plásmido de las bibliotecas en la reacción de transformación para obtener >10⁶ UFC. Tanto la NGA como el ARNgu en la cadena principal de pRSF-1b estaban bajo el control de promotores de T7. Se dejó que la reacción de transformación se recuperara durante 1 hora, después de lo cual se diluyó en medio LB que contenía carbenicilina y kanamicina y se cultivó durante una noche. Al día siguiente, la mezcla se diluyó en medio autoinductor Overnight Express™ Instant TB (Millipore Sigma) para permitir la expresión de la NGA y del ARNgu, y se cultivó durante 4 h o 20 h adicionales, después de lo cual las células se centrifugaron y se aisló el ADN plasmídico con un kit de minipreparación (Qiagen, Germantown, MD). En presencia del ARNgu apropiado, los plásmidos conteniendo un PAM que es reconocible por la NGA se escindirán, lo que dará como resultado su eliminación de la población. Los plásmidos que contienen PAM que no son reconocibles por la NGA o que se transforman en bacterias que no contienen un ARNgu apropiado, sobrevivirán y se replicarán. Las regiones PAM y protoespaciadoras de plásmidos no escindidos se amplificaron por PCR y se prepararon para la secuenciación siguiendo los protocolos publicados (guía de preparación de bibliotecas metagenómicas de 16s 15044223B, Illumina, San Diego, CA). La secuenciación profunda (lecturas de extremo único de 75 pb) se realizó en un MiSeq (Illumina) por un proveedor de servicios (MoGene, St. Louis, MO). Típicamente, se obtuvieron lecturas de 1-4 M por amplicón. Se extrajeron las regiones PAM, se contaron y se normalizaron con respecto a las lecturas totales para cada muestra. Los PAM que conducen a la escisión del plásmido se identificaron por estar subrepresentados en comparación con los controles (es decir, cuando la biblioteca se transforma en *E. coli* que contiene la NGA pero carece de un ARNgu apropiado). Para representar los requisitos de PAM para una NGA novedosa, las relaciones de empobrecimiento (frecuencia en la muestra/frecuencia en el control) para todas las secuencias en la región en cuestión se convirtieron en valores de enriquecimiento con una transformación de -log base 2. Se definieron PAM suficientes como los que tenían valores de enriquecimiento >2,3 (lo que corresponde a índices de empobrecimiento de < ~0,2). Los PAM por encima de este umbral en ambas bibliotecas se recopilaron y utilizaron para generar anagramas web logo, que, por ejemplo, se puede generar utilizando un servicio en Internet conocido como "weblogo". Las secuencias de PAM se identificaron e informaron cuando había un patrón constante en los PAM más enriquecidos. En la Tabla 2 se proporciona una secuencia PAM de consenso (que tiene un factor de enriquecimiento (FE) >2,3) para cada NGA. La orientación de la PAM también se indica en la Tabla 2.

Tabla 2: Determinación PAM o similar a PAM

ID de NGA	ARNgu de L1 (SEQ ID NO:)	ARNgu de L2 (SEQ ID NO:)	PAM (SEQ ID NO:)	Orientación PAM
APG06622	5	6	7	5'-diana-PAM-3'
APG02787	12	13	14	5'-diana-PAM-3'
APG06248	19	20	21	5'-diana-PAM-3'
APG06007	26	27	28	5'-diana-PAM-3'
APG02874	33	34	35	5'-diana-PAM-3'
APG03850	40	41	42	5'-diana-PAM-3'
APG07553	47	48	49	5'-diana-PAM-3'
APG03031	54	55	35	5'-diana-PAM-3'
APG09208	60	61	62	5'-diana-PAM-3'
APG05586	67	68	69	5'-diana-PAM-3'
APG08770	74	75	69	5'-diana-PAM-3'
APG08167	80	81	82	5'-diana-PAM-3'
APG01604	87	88	82	5'-diana-PAM-3'
APG03021	93	94	95	5'-diana-PAM-3'
APG06015	100	101	102	5'-diana-PAM-3'
APG09344	107	108	109	5'-diana-PAM-3'
APG07991	114	115	116	5'-diana-PAM-3'
APG01868	121	122	116	5'-diana-PAM-3'
APG02998	127	128	116	5'-diana-PAM-3'

Ejemplo 4: Demostración de la actividad de edición génica en células de mamífero

Se produjeron casetes de expresión de NGA y se introdujeron en vectores para la expresión en mamífero. Cada NGA se optimizó mediante codones para la expresión humana (SEQ ID NO: 134-152) y se fusionó operativamente en el extremo 5' a una secuencia de localización nuclear de SV40 (NLS; SEQ ID NO: 251) y a 3 etiquetas FLAG (SEQ ID NO: 252) y se fusionaron operativamente en el extremo 3' a secuencias SLN de nucleoplasmina (SEQ ID NO: 253). Se utilizaron dos copias de la secuencia SLN, fusionados operativamente en tándem. Cada casete de expresión estaba bajo el control de un promotor de citomegalovirus (CMV) (SEQ ID NO: 258). Se sabe en la técnica que el potenciador de la transcripción de CMB (SEQ ID NO: 259) también puede incluirse en construcciones que comprenden el promotor de CMV. Se produjeron construcciones de expresión de ARN de guía codificando un ARNg único, cada una bajo el control de un promotor U6 de la ARN polimerasa III humana (SEQ ID NO: 260) y se introdujeron en el vector pTwist High Copy Amp. Las secuencias para las secuencias diana de cada guía figuran en la Tabla 3.

Varias de las construcciones descritas anteriormente se introdujeron en células de mamífero. Un día antes de la transfección, se sembraron 1×10^5 células HEK293T (Sigma) en placas de 24 pocillos en medio Eagle modificado por Dulbecco (DMEM) más suero bovino fetal al 10 % (vol/vol) (Gibco) y penicilina-estreptomycin al 1 % (Gibco). Al día siguiente, cuando las células tenían una confluencia del 50-60 %, se cotransfectaron 500 ng de un plásmido de expresión de NGA más 500 ng de un plásmido de expresión de ARNg único utilizando 1,5 μ l de Lipofectamine 3000 (Thermo Scientific) por pocillo, siguiendo las instrucciones del fabricante. Después de 48 horas de cultivo, el ADN genómico total se recogió utilizando un kit de aislamiento de ADN genómico (Machery-Nagel) de acuerdo con las instrucciones del fabricante.

A continuación, se analizó el ADN genómico total para determinar la tasa de edición de cada NGA para cada diana genómica. En primer lugar, se produjeron oligonucleótidos para usarse para amplificación por PCR y el análisis posterior del sitio diana genómico amplificado. Las secuencias de oligonucleótidos utilizadas se enumeran en la Tabla 4.

Todas las reacciones de PCR se realizaron utilizando 10 μ l de ADN polimerasa 2X Master Mix Phusion High-Fidelity (Thermo Scientific) en una reacción de 20 μ l que incluía 0,5 μ M de cada cebador. Primero se amplificaron grandes

regiones genómicas abarcando cada gen diana usando los cebadores de PCR n.º 1, utilizando un programa de: 98 °C, 1 min; 30 ciclos de [98 °C, 10 s; 62 °C, 15 s; 72 °C, 5 min]; 72 °C, 5 min; 12 °C, mantenido. Luego se amplificó adicionalmente un microlitro de esta reacción de PCR utilizando cebadores específicos para cada guía (cebadores de PCR n.º 2) usando un programa de: 98 °C, 1 min; 35 ciclos de [98 °C, 10 s; 67 °C, 15 s; 72 °C, 30 s]; 72 °C, 5 min; 12 °C, mantenido. Los cebadores para PCR n.º 2 incluyen las secuencias protuberantes Nextera Read 1 and Read 2 Transposase Adapter para la secuenciación por Illumina.

Para las NGA APG02874, APG03850 y APG09208, los métodos se llevaron a cabo como se describe anteriormente. Se seleccionaron varios genes diferentes del genoma humano para realizar la escisión guiada por ARN. Estos locus se incluyen en la siguiente Tabla 3, junto con la referencia a la SEQ ID NO del ARN_g. El porcentaje de indel, que es una indicación de la actividad de la NGA, también se muestra.

Tabla 3: Secuencias diana y ARN_g de los ARN guía utilizados para probar la actividad de edición génica en células de mamífero

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARN _g (SEQ ID NO:)
APG02874, APG09208	RelA	SGN000973, SGN000778	153	188, 206
APG02874	RelA	SGN000974	154	189
APG02874	RelA	SGN000975, SGN000780	155	190, 208
APG02874, APG09208	AurkB	SGN000976, SGN000775	156	191, 204
APG02874, APG09208	AurkB	SGN000977, SGN000776	157	192, 205
APG02874	AurkB	SGN000978	158	193
APG02874	VEGFA	SGN000979	159	194
APG02874	VEGFA	SGN000981	160	195
APG03850	RelA	SGN000982	161	196
APG03850	RelA	SGN000983	162	197
APG03850	RelA	SGN000984	163	198
APG03850	AurkB	SGN000985	164	199
APG03850	AurkB	SGN000986	165	200
APG03850	AurkB	SGN000987	166	201
APG03850	VEGFA	SGN000988	167	202
APG03850	VEGFA	SGN000990	168	203
APG09208	RelA	SGN000779	169	207
APG09208	AurkB	SGN000793	170	209
APG09208	AurkB	SGN000794	171	210
APG05586	TRA	SGN001163	553	557
APG05586	TRA	SGN001164	554	558
APG05586	VEGFA	SGN001165	555	559
APG05586	VEGFA	SGN001166	556	560
APG09208	RelA	SGN000778	153	206
APG09208	AurkB	SGN000793	609	811
APG05586, APG08770, APG09298	EMX1	SGN001159	610	812
APG05586, APG08770, APG09298	TRA	SGN001162	611	813

ES 3 029 182 T3

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG05586, APG08770, APG09298	TRA	SGN001163	612	814
APG05586, APG08770, APG09298	TRA	SGN001164	613	815
APG05586, APG08770, APG09298	VEGFA	SGN001165	614	816
APG05586, APG08770, APG09298	VEGFA	SGN001166	615	817
APG05586, APG09298	VEGFA	SGN001167	616	818
APG09208	RelA	SGN001213	617	819
APG08167, APG01604	VEGFA	SGN001245	618	820
APG08167, APG01604	VEGFA	SGN001246	619	821
APG08167, APG01604	VEGFA	SGN001247	620	822
APG08167, APG01604	RelA	SGN001248	621	823
APG08167, APG01604	RelA	SGN001249	622	824
APG08167, APG01604	RelA	SGN001250	623	825
APG08167, APG01604	AurkB	SGN001251	624	826
APG08167, APG01604	AurkB	SGN001252	625	827
APG08167, APG01604	AurkB	SGN001253	626	828
APG07991	RelA	SGN001312	627	829
APG07991	RelA	SGN001313	628	830
APG07991	RelA	SGN001314	629	831
APG01868	RelA	SGN001315	630	832
APG01868	RelA	SGN001316	631	833
APG01868	RelA	SGN001317	632	834
APG02998	RelA	SGN001318	633	835
APG02998	RelA	SGN001319	634	836
APG02998	RelA	SGN001320	635	837
APG09344	RelA	SGN001321	636	838
APG06015	RelA	SGN001322	637	839
APG03021	RelA	SGN001323	638	840
APG09344	RelA	SGN001324	639	841
APG06015	RelA	SGN001325	640	842
APG03021	RelA	SGN001326	641	843
APG06015	RelA	SGN001327	642	844
APG03021	RelA	SGN001328	643	845
APG09344	RelA	SGN001329	644	846
APG03021	TRA	SGN001330	645	847
APG03021	TRA	SGN001331	646	848
APG03021	TRA	SGN001332	647	849

ES 3 029 182 T3

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG06015	TRA	SGN001333	648	850
APG06015	TRA	SGN001334	649	851
APG06015	TRA	SGN001335	650	852
APG09344	TRA	SGN001336	651	853
APG09344	TRA	SGN001337	652	854
APG09344	TRA	SGN001338	653	855
APG07991	TRA	SGN001339	654	856
APG07991	TRA	SGN001340	655	857
APG07991	TRA	SGN001341	656	858
APG01868	TRA	SGN001342	657	859
APG01868	TRA	SGN001343	658	860
APG01868	TRA	SGN001344	659	861
APG01868	TRA	SGN001692	660	862
APG02998	TRA	SGN001345	661	863
APG02998	TRA	SGN001346	662	864
APG02998	TRA	SGN001347	663	865
APG03021	HAO1	SGN001348	664	866
APG03021	HAO1	SGN001349	665	867
APG03021	HAO1	SGN001350	666	868
APG06015	HAO1	SGN001351	667	869
APG06015	HAO1	SGN001352	668	870
APG06015	HAO1	SGN001353	669	871
APG09344	HAO1	SGN001354	670	872
APG09344	HAO1	SGN001355	671	873
APG09344	HAO1	SGN001356	672	874
APG07991	HAO1	SGN001357	673	875
APG07991	HAO1	SGN001358	674	876
APG07991	HAO1	SGN001359	675	877
APG01868	HAO1	SGN001360	676	878
APG01868	HAO1	SGN001361	677	879
APG01868	HAO1	SGN001362	678	880
APG02998	HAO1	SGN001363	679	881
APG02998	HAO1	SGN001364	680	882
APG02998	HAO1	SGN001365	681	883
APG05586, APG09298	TRA	SGN001371	682	884
APG05586, APG09298	TRA	SGN001372	683	885

ES 3 029 182 T3

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG05586, APG09298	TRA	SGN001373	684	886
APG05586, APG09298	TRA	SGN001374	685	887
APG05586, APG09298	TRA	SGN001375	686	888
APG05586, APG09298	TRA	SGN001376	687	889
APG05586, APG09298	TRA	SGN001377	688	890
APG05586, APG09298	TRA	SGN001378	689	891
APG05586, APG09298	TRA	SGN001379	690	892
APG05586, APG09298	TRA	SGN001380	691	893
APG05586, APG09298	TRA	SGN001381	692	894
APG05586, APG09298	TRA	SGN001382	693	895
APG05586, APG09298	B2M	SGN001383	694	896
APG05586, APG09298	B2M	SGN001384	695	897
APG05586, APG09298	B2M	SGN001385	696	898
APG05586, APG09298	B2M	SGN001386	697	899
APG05586, APG09298	B2M	SGN001387	698	900
APG05586, APG09298	B2M	SGN001388	699	901
APG05586, APG09298	B2M	SGN001389	700	902
APG05586, APG09298	B2M	SGN001390	701	903
APG05586, APG09298	B2M	SGN001391	702	904
APG05586, APG09298	B2M	SGN001392	703	905
APG05586, APG09298	B2M	SGN001393	704	906
APG05586, APG09298	B2M	SGN001394	705	907
APG05586, APG09298	LDHA	SGN001395	706	908
APG05586, APG09298	LDHA	SGN001396	707	909
APG05586, APG09298	LDHA	SGN001397	708	910
APG05586, APG09298	LDHA	SGN001399	709	911
APG05586, APG09298	LDHA	SGN001400	710	912
APG05586, APG09298	LDHA	SGN001401	711	913
APG05586, APG09298	LDHA	SGN001402	712	914
APG05586, APG09298	LDHA	SGN001403	713	915
APG05586, APG09298	LDHA	SGN001404	714	916
APG05586, APG09298	LDHA	SGN001405	715	917
APG05586, APG09298	HAO1	SGN001406	716	918
APG05586, APG09298	HAO1	SGN001407	717	919
APG05586, APG09298	HAO1	SGN001408	718	920
APG05586, APG09298	HAO1	SGN001409	719	921

ES 3 029 182 T3

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG05586, APG09298	HAO1	SGN001410	720	922
APG05586, APG09298	HAO1	SGN001411	721	923
APG05586, APG09298	HAO1	SGN001412	722	924
APG05586, APG09298	HAO1	SGN001413	723	925
APG05586, APG09298	HAO1	SGN001414	724	926
APG05586, APG09298	HAO1	SGN001415	725	927
APG05586, APG09298	HAO1	SGN001416	726	928
APG01604	B2M	SGN001592	727	929
APG01604	B2M	SGN001593	728	930
APG01604	B2M	SGN001594	729	931
APG01604	B2M	SGN001595	730	932
APG01604	B2M	SGN001596	731	933
APG01604	B2M	SGN001597	732	934
APG01604	B2M	SGN001598	733	935
APG01604	B2M	SGN001599	734	936
APG01604	B2M	SGN001600	735	937
APG01604	B2M	SGN001601	736	938
APG01604	B2M	SGN001602	737	939
APG01604	B2M	SGN001603	738	940
APG01604	HAO1	SGN001616	739	941
APG01604	HAO1	SGN001617	740	942
APG01604	HAO1	SGN001618	741	943
APG01604	HAO1	SGN001619	742	944
APG01604	HAO1	SGN001620	743	945
APG01604	HAO1	SGN001621	744	946
APG01604	HAO1	SGN001622	745	947
APG01604	HAO1	SGN001623	746	948
APG01604	HAO1	SGN001624	747	949
APG01604	HAO1	SGN001625	748	950
APG01604	HAO1	SGN001626	749	951
APG01604	HAO1	SGN001627	750	952
APG01604	LDHA	SGN001640	751	953
APG01604	LDHA	SGN001641	752	954
APG01604	LDHA	SGN001642	753	955
APG01604	LDHA	SGN001643	754	956
APG01604	LDHA	SGN001644	755	957

ES 3 029 182 T3

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG01604	LDHA	SGN001645	756	958
APG01604	LDHA	SGN001646	757	959
APG01604	LDHA	SGN001647	758	960
APG01604	LDHA	SGN001648	759	961
APG01604	LDHA	SGN001649	760	962
APG01604	LDHA	SGN001650	761	963
APG01604	LDHA	SGN001651	762	964
APG01604	TRA	SGN001664	763	965
APG01604	TRA	SGN001665	764	966
APG01604	TRA	SGN001666	765	967
APG01604	TRA	SGN001667	766	968
APG01604	TRA	SGN001668	767	969
APG01604	TRA	SGN001669	768	970
APG01604	TRA	SGN001670	769	971
APG01604	TRA	SGN001671	770	972
APG01604	TRA	SGN001672	771	973
APG01604	TRA	SGN001673	772	974
APG01604	TRA	SGN001674	773	975
APG01604	TRA	SGN001675	774	976
APG01868	TRA	SGN001684	775	977
APG01868	TRA	SGN001685	776	978
APG01868	TRA	SGN001686	777	979
APG01868	TRA	SGN001687	778	980
APG01868	TRA	SGN001688	779	981
APG01868	TRA	SGN001689	780	982
APG01868	TRA	SGN001690	781	983
APG01868	TRA	SGN001691	782	984
APG01868	EMX1	SGN001697	783	985
APG01868	EMX1	SGN001698	784	986
APG01868	EMX1	SGN001699	785	987
APG01868	EMX1	SGN001700	786	988
APG01868	EMX1	SGN001701	787	989
APG01868	EMX1	SGN001702	788	990
APG01868	EMX1	SGN001703	789	991
APG01868	EMX1	SGN001704	1176	992
APG01868	EMX1	SGN001705	1177	993

(continuación)

ID de NGA	Gen	ID de la Guía	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)
APG01868	EMX1	SGN001706	1178	994
APG01868	EMX1	SGN001707	1179	995
APG01868	EMX1	SGN001708	1180	996
APG01868	EMX1	SGN001709	1181	997
APG01868	EMX1	SGN001710	1182	998
APG01868	EMX1	SGN001711	1183	999
APG01868	HAO1	SGN001713	1184	1000
APG01868	HAO1	SGN001714	1185	1001
APG01868	HAO1	SGN001715	790	1002
APG01868	HAO1	SGN001716	791	1003
APG01868	HAO1	SGN001717	792	1004
APG01868	HAO1	SGN001718	793	1005
APG01868	HAO1	SGN001719	794	1006
APG01868	HAO1	SGN001721	795	1007
APG01868	HAO1	SGN001722	796	1008
APG01868	HAO1	SGN001723	797	1009
APG01868	HAO1	SGN001724	798	1010
APG01868	LDHA	SGN001725	799	1011
APG01868	LDHA	SGN001726	800	1012
APG01868	LDHA	SGN001727	801	1013
APG01868	LDHA	SGN001728	802	1014
APG01868	LDHA	SGN001729	803	1015
APG01868	LDHA	SGN001730	804	1016
APG01868	LDHA	SGN001731	805	1017
APG01868	LDHA	SGN001732	806	1018
APG01868	LDHA	SGN001733	807	1019
APG01868	LDHA	SGN001734	808	1020
APG01868	LDHA	SGN001735	809	1021
APG01868	VEGFA	SGN001785	810	1022

Tabla 4: Oligonucleótidos para la detección de la actividad de edición génica en células de mamífero

Descripción	Secuencia de Cebador	SEQ ID NO
SGN000977 DIR	CTTGCTAGCTGGAGGTCCATC	172
SGN000977 INV	TGTTGGCAAATCTAGTCTCG	173
SGN000978 DIR	ACATTTGACGAGCAGCGAA	174

(continuación)

Descripción	Secuencia de Cebador	SEQ ID NO
SGN000978 INV	GGCCCCCTGGAGAGGTTTTAA	175
SGN000979, SGN000982, SGN000988, SGN000990 DIR	ACACAGCTTCCCGTTCTCAG	176
SGN000979, SGN000982, SGN000988, SGN000990 INV	ATTCACCCAGCTTCCCTGTG	177
SGN000981 DIR	GGCGTCGCACTGAAACTTTT	178
SGN000981 INV	AGTTCATGGTTTCGGAGGCC	179
SGN000983 DIR	CGACCAAACAAGTGCAAAGG	180
SGN000983 INV	GGGTTGTTGTTGGTCTGGAT	181
SGN000775, SGN000776, SGN000793, SGN000794, SGN000976, SGN000985, SGN000986, SGN000987 DIR	ACTGCCATGGGAAGAAGGTG	182
SGN000775, SGN000776, SGN000793, SGN000794, SGN000976, SGN000985, SGN000986, SGN000987 INV	ACAATTCTCCTGCCTCAGCC	183
SGN000778, SGN000973, SGN000984 DIR	TGGCCCCCTATGTGGAGATCA	184
SGN000778, SGN000973, SGN000984 INV	GGCAGAGCTCAGCCTCATAG	185
SGN000779, SGN000780, SGN000974, SGN000975 DIR	ATATCCCCACTTCCCCTGCT	186
SGN000779, SGN000780, SGN000974, SGN000975 INV	CACCTCAAGGACAGCTCTGG	187
SGN001163, SGN001164 DIR	TTGATAGCTTGTGCCTGTCC	561
SGN001163, SGN001164 INV	AGAGTCTCTCAGCTGGTACA	562
SGN001165, SGN001166 DIR	GCGACAGGGGCAAAGTGAGT	563
SGN001165, SGN001166 INV	CTAGCACTTCTCGCGGCTCC	564
SGN001382 DIR	AACTCATGCCTGCTGCTCTT	1023
SGN001382 INV	CAGTCTCACGCAGTCACTCA	1024
SGN001371, SGN001372, SGN001373, SGN001689, SGN001690 DIR	AACTGAGGCGGCTGAAATGA	1025
SGN001371, SGN001372, SGN001373, SGN001689, SGN001690 INV	TGGGACATGCAAGCCCATAA	1026
SGN001730 DIR	AAGATGTTGACATGCTCTTCC	1027
SGN001730 INV	TATGCAGTCAAAAGCCTCA	1028
SGN001353, SGN001356, SGN001359, SGN001362, SGN001365, SGN001406, SGN001407, SGN001412, SGN001413, SGN001414, SGN001622, SGN001623, SGN001624, SGN001625, SGN001626, SGN001627 DIR	AAGTCATTGCTTGTTTGA	1029
SGN001353, SGN001356, SGN001359, SGN001362, SGN001365, SGN001406, SGN001407, SGN001412, SGN001413, SGN001414, SGN001622, SGN001623, SGN001624, SGN001625, SGN001626, SGN001627 INV	TGGTGCATTGAGAGAAGGAG	1030
SGN001375, SGN001376, SGN001377 DIR	AATGAAGCCAGGCAAGAGCA	1031
SGN001375, SGN001376, SGN001377 INV	CTGTGCAAACCCAGGCTAGA	1032
SGN001251, SGN001252 DIR	ACATTTGACGAGCAGCGAA	1033
SGN001251, SGN001252 INV	GGCCCCCTGGAGAGGTTTTAA	1034
SGN001380 DIR	ACCCGGCCTGCTTTTCTTAA	1035
SGN001380 INV	GGCAGCGAGGCATACATAGT	1036
SGN001731, SGN001732, SGN001733, SGN001734, SGN001735 DIR	ACCCTGCTTTTCTGCCTTT	1037

(continuación)

Descripción	Secuencia de Cebador	SEQ ID NO
SGN001731, SGN001732, SGN001733, SGN001734, SGN001735 INV	CAGGCCTAATGGACATTAATCCT	1038
SGN001374,SGN001688 DIR	ACTCACTAAGGGGCCCCATCT	1039
SGN001374,SGN001688 INV	CAGGAGGAGGATTCGGAACC	1040
SGN001726,SGN001727,SGN001728,SGN001729 DIR	AGGAAAATGAATCACAATTACT	1041
SGN001726,SGN001727,SGN001728,SGN001729 INV	GTGCGAAAGGGCAAGATTCT	1042
SGN001397 DIR	AGGCCTTTCAACTCTCTTTTGGCA	1043
SGN001397 INV	GGATGGGGTCAAGGTATGGGC	1044
SGN001213, SGN001248, SGN001249, SGN001321, SGN001322,SGN001323 DIR	ATGACATTCAGGCCACAGTG	1045
SGN001213, SGN001248, SGN001249, SGN001321, SGN001322,SGN001323 INV	CTTCCTCCTATTTCAGGCCCA	1046
SGN001725 DIR	CAGCTTTTCAAATGGGGTGC	1047
SGN001725 INV	CAACAAATGGAGACCATCTGGA	1048
SGN001381 DIR	CAGTATTCTAAGGACGCCAGAA	1049
SGN001381 INV	GCACTTTGGGAGGCTGAA	1050
SGN001390, SGN001391, SGN001392, SGN001393, SGN001394, SGN001592, SGN001593, SGN001594, SGN001595, SGN001596, SGN001597, SGN001598, SGN001599, SGN001600, SGN001601, SGN001602, SGN001603 DIR	CGGGCATTCTGAAGCTG	1051
SGN001390, SGN001391, SGN001392, SGN001393, SGN001394, SGN001592, SGN001593, SGN001594, SGN001595, SGN001596, SGN001597, SGN001598, SGN001599, SGN001600, SGN001601, SGN001602, SGN001603 INV	GTAGGCCAAAGGTCTCCCC	1052
SGN001383, SGN001384, SGN001385, SGN001386, SGN001387,SGN001388,SGN001389 DIR	CTTGACACCAAGTTAGCCCC	1053
SGN001383, SGN001384, SGN001385, SGN001386, SGN001387,SGN001388,SGN001389 INV	TCATACACAACTTTCAGCAGC	1054
SGN000793,SGN001253 DIR	CTTGAGCTGGAGGTCCATC	1055
SGN000793,SGN001253 INV	TGTTGGCAAATCTAGTCTCG	1056
SGN001379 DIR	GAGCAGCTGAGTCAATGATAGT	1057
SGN001379 INV	GGAGAGATCTGGAGGGAACCTTA	1058
SGN001165,SGN001166,SGN001245 DIR	GCAAAGTGAGTGACCTGCTT	1059
SGN001165,SGN001166,SGN001245 INV	GAGCTAGCACTTCTCGCG	1060
SGN001738 DIR	GCTGTTTGGGAGGTCAGAAA	1061
SGN001738 INV	GAATATTGAAGGGGGCAGGG	1062
SGN001167,SGN001246,SGN001247,SGN001785 DIR	GGACACTTCCCAAAGGACC	1063
SGN001167,SGN001246,SGN001247,SGN001785 INV	CACGTCCTCACTCTCGAAGA	1064
SGN001399, SGN001402, SGN001403, SGN001646, SGN001647, SGN001648, SGN001649, SGN001650, SGN001651 DIR	GGCCTTCACTCTTCACAGACCC	1065
SGN001399, SGN001402, SGN001403, SGN001646, SGN001647, SGN001648, SGN001649, SGN001650, SGN001651 INV	GGATGGGGTCAAGGTATGGGC	1066
SGN001162, SGN001163, SGN001164, SGN001330, SGN001331, SGN001332, SGN001333, SGN001334, SGN001335, SGN001336, SGN001337, SGN001338, SGN001339, SGN001340, SGN001341,	GGGCAAAGAGGGAAATGAGA	1067

(continuación)

Descripción	Secuencia de Cebador	SEQ ID NO
SGN001342, SGN001343, SGN001344, SGN001345, SGN001346, SGN001347, SGN001664, SGN001665, SGN001666, SGN001667, SGN001668, SGN001669, SGN001670, SGN001671, SGN001672, SGN001673, SGN001674, SGN001675, SGN001684, SGN001685, SGN001686, SGN001687, SGN001691, SGN001692 DIR		
SGN001162, SGN001163, SGN001164, SGN001330, SGN001331, SGN001332, SGN001333, SGN001334, SGN001335, SGN001336, SGN001337, SGN001338, SGN001339, SGN001340, SGN001341, SGN001342, SGN001343, SGN001344, SGN001345, SGN001346, SGN001347, SGN001664, SGN001665, SGN001666, SGN001667, SGN001668, SGN001669, SGN001670, SGN001671, SGN001672, SGN001673, SGN001674, SGN001675, SGN001684, SGN001685, SGN001686, SGN001687, SGN001691, SGN001692 INV	GAACCTGGCCATTCTGAAG	1068
SGN001714, SGN001722 DIR	GGTTTTTGAGGTGGAGTTGA	1069
SGN001714, SGN001722 INV	CCCCCTAACCAAGTGAAAAGA	1070
SGN001716, SGN001717, SGN001721 DIR	TAGATAAATGAGCAGTGAACA GCC	1071
SGN001716, SGN001717, SGN001721 INV	TCCACAAAGGATCACAAAGTC A	1072
SGN001313, SGN001316, SGN001319 DIR	TGAGAGACAGTGGGACAGAC	1073
SGN001313, SGN001316, SGN001319 INV	AGTCCTAGAGGAGGCAGAAC	1074
SGN001702, SGN001703, SGN001707, SGN001709, SGN001710, SGN001711 DIR	TGAGTCCGAGCAGAAGAAGA	1075
SGN001702, SGN001703, SGN001707, SGN001709, SGN001710, SGN001711 INV	GGAGATTGGAGACACGGAGA	1076
SGN001723 DIR	TGGAGGTGGAGTTGAATAACA	1077
SGN001723 INV	CTTCTCCCCCTAACCAAGTG	1078
SGN001395, SGN001396 DIR	TGGATCTCCAACATGGCAGCC	1079
SGN001395, SGN001396 INV	CGGAAGGCTAAGGAGGGAGG A	1080
SGN000778, SGN001250, SGN001312, SGN001314, SGN001315, SGN001317, SGN001318, SGN001320, SGN001324, SGN001325, SGN001326, SGN001327, SGN001328, SGN001329 DIR	TGGCCCCTATGTGGAGATCA	1081
SGN000778, SGN001250, SGN001312, SGN001314, SGN001315, SGN001317, SGN001318, SGN001320, SGN001324, SGN001325, SGN001326, SGN001327, SGN001328, SGN001329 INV	GGCAGAGCTCAGCCTCATAG	1082
SGN001159, SGN001697, SGN001698, SGN001699, SGN001700, SGN001701, SGN001704, SGN001705, SGN001706, SGN001708 DIR	TGTTAGACCCATGGGAGCAG	1083
SGN001159, SGN001697, SGN001698, SGN001699, SGN001700, SGN001701, SGN001704, SGN001705, SGN001706, SGN001708 INV	GTTGCCACCCCTAGTCATT	1084
SGN001400, SGN001401, SGN001404, SGN001405, SGN001640, SGN001641, SGN001642, SGN001643, SGN001644, SGN001645 DIR	TTCCACGCTAAGGTATGGGCC	1085
SGN001400, SGN001401, SGN001404, SGN001405, SGN001640, SGN001641, SGN001642, SGN001643, SGN001644, SGN001645 INV	GCCAACAGCACCAACCCCAA	1086
SGN001348, SGN001349, SGN001350, SGN001351, SGN001352, SGN001354, SGN001355, SGN001357, SGN001358, SGN001360, SGN001361, SGN001363, SGN001364, SGN001408, SGN001409, SGN001410, SGN001411, SGN001415, SGN001416, SGN001616, SGN001617, SGN001618, SGN001619, SGN001620, SGN001621, SGN001713, SGN001715, SGN001718, SGN001719, SGN001724 DIR	TTGCTCACTTGATGTAAGCAA	1087

(continuación)

Descripción	Secuencia de Cebador	SEQ ID NO
SGN001348, SGN001349, SGN001350, SGN001351, SGN001352, SGN001354, SGN001355, SGN001357, SGN001358, SGN001360, SGN001361, SGN001363, SGN001364, SGN001408, SGN001409, SGN001410, SGN001411, SGN001415, SGN001416, SGN001616, SGN001617, SGN001618, SGN001619, SGN001620, SGN001621, SGN001713, SGN001715, SGN001718, SGN001719, SGN001724 INV	TTTTGGTACGGTCTTTGTGT	1088

El ADN genómico purificado se sometió a PCR N.º 1 y PCR N.º 2 como se ha indicado anteriormente. Después de la segunda amplificación por PCR, el ADN se limpió utilizando un kit de limpieza de PCR (Zymo) de acuerdo con las instrucciones del fabricante y se eluyó en agua. Se combinaron 200-500 ng de producto de PCR n.º 2 purificado con 2 µl de tampón NEB 2 10X y agua en una reacción de 20 µl y se hibridaron para formar heterodúplex de ADN usando un programa de: 95 °C, 5 min; 95-85 °C, enfriado a una velocidad de 2 °C/s; 85-25 °C, enfriado a una velocidad de 0,1 °C/s; 12 °C, mantenido. Tras la hibridación se retiraron 5 µl de ADN como control sin enzima, se añadió 1 µl de endonucleasa I de T7 (NEB) y la reacción se incubó a 37 °C durante 1 hora. Tras la incubación se añadió colorante de carga FlashGel 5x (Lonza) y se analizaron 5 µl de cada reacción y los controles mediante un FlashGel de agarosa al 2,2 % (Lonza) mediante electroforesis en gel. Después de la visualización del gel, se determinó el porcentaje de unión de extremos no homólogos (NHEJ) mediante la siguiente ecuación: % de sucesos de NHEJ = $100 \times [1 - (1 - \text{fracción escindida})^{1/2}]$, donde (fracción escindida) se define como: (densidad de productos digeridos)/(densidad de productos digeridos + banda parental no digerida).

Para algunas muestras, se utilizó SURVEYOR® para analizar los resultados después de la expresión en células de mamífero. Las células se incubaron a 37 °C durante 72 h postransfección antes de la extracción de ADN genómico. El ADN genómico se extrajo utilizando la solución de extracción de ADN QuickExtract (Epicentre) siguiendo el protocolo del fabricante. La región genómica que flanquea el sitio diana de la NGA se amplificó por PCR y los productos se purificaron utilizando la columna QiaQuick Spin (Qiagen) siguiendo el protocolo del fabricante. Se mezclaron 200-500 ng totales de los productos de PCR purificados con 1 µl de tampón Taq DNA Polymerase PCR 10× (Enzymatics) y agua ultrapura hasta un volumen final de 10 µl, y se somete a un proceso de rehibridación para permitir la formación del heterodúplex: 95 °C durante 10 min, 95 °C a 85 °C aumentando a -2 °C/s, 85 °C a 25 °C a -0,25 °C/s, y 25 °C mantenidos durante 1 min.

Después de la rehibridación, los productos se trataron con nucleasa SURVEYOR® y potenciador S SURVEYOR® (Integrated DNA Technologies) siguiendo el protocolo recomendado por el fabricante y se analizan en geles de poliacrilamida Novex TBE al 4-20 % (Life Technologies). Los geles se tiñeron con tinción de ADN SYBR Gold (Life Technologies) durante 10 min y se tomaron imágenes con un Gel Doc gel imaging system (Bio-rad). La cuantificación se basó en intensidades de banda relativas. El porcentaje de indeles se determinó mediante la fórmula, $100 \times (1 - ((b+c)/(a+b+c))^{1/2})$, en que a es la intensidad integrada del producto de PCR no digerido, y b y c son las intensidades integradas de cada producto de escisión.

Además, los productos de PCR n.º 2 conteniendo secuencias protuberantes de Illumina se sometieron a la preparación de la biblioteca siguiendo el protocolo de bibliotecas de secuenciación metagenómica de Illumina de 16S. Un proveedor de servicios (MOGene) realizó la secuenciación profunda en una plataforma Illumina Mi-Seq. Típicamente, se generan 200.000 lecturas de extremos emparejados de 250 pb (2 x 100.000 lecturas) por amplicón. Las lecturas se analizaron utilizando CRISPResso (Pinello, *et al.* 2016 Nature Biotech, 34:695-697) para calcular las tasas de edición. Los alineamientos de salida se seleccionaron manualmente para confirmar los sitios de inserción y delección, así como para identificar sitios de microhomología en los sitios de recombinación. Las tasas de edición se muestran en la Tabla 5. Todos los experimentos se realizaron en células humanas. La "secuencia diana" es la secuencia que es objetivo dentro del gen diana. Para cada secuencia diana, el ARN de guía comprendía la secuencia diana del ARN complementario y el ARN_g apropiado de acuerdo con la NGA utilizada. En las Tablas 6.1-6.3 se muestra un desglose seleccionado de experimentos por ARN de guía.

Tabla 5: Actividad de las NGA en células de mamífero

NGA	ID de la Guía	Diana génica	Tasa global de edición	Tasa de delección en la muestra	Tasa de inserción en la muestra
APG02874	SGN000973	RelA	N.D.		
APG02874	SGN000974	RelA	0,10 %	37,23 %	62,77 %
APG02874	SGN000975	RelA	N.D.		

(continuación)

NGA	ID de la Guía	Diana génica	Tasa global de edición	Tasa de delección en la muestra	Tasa de inserción en la muestra
APG02874	SGN000976	AurkB	0,56 %	35,17 %	64,84 %
APG02874	SGN000977	AurkB	0,30 %	76,34 %	23,65 %
APG02874	SGN000978	AurkB	N.D.		
APG02874	SGN000979	VEGFA	0,17 %	13,95 %	86,05 %
APG02874	SGN000981	VEGFA	N.D.		
APG03850	SGN000982	RelA	0,19 %	40,55 %	59,45 %
APG03850	SGN000983	RelA	N.D.		
APG03850	SGN000984	RelA	0,05 %	72,17 %	27,84 %
APG03850	SGN000985	AurkB	0,37 %	66,62 %	33,37 %
APG03850	SGN000986	AurkB	N.D.		
APG03850	SGN000987	AurkB	N.D.		
APG03850	SGN000988	VEGFA	N.D.		
APG03850	SGN000990	VEGFA	N.D.		
APG09208	SGN000775	AurkB	N.D.		
APG09208	SGN000776	AurkB	N.D.		
APG09208	SGN000778	RelA	N.D.		
APG09208	SGN000779	RelA	N.D.		
APG09208	SGN000780	RelA	N.D.		
APG09208	SGN000793	AurkB	1,18 %	59,97 %	40,02 %
APG09208	SGN000794	AurkB	N.D.		
APG05586	SGN001163	TRA	23,49 %	95,03 %	5,41 %
APG05586	SGN001164	TRA	1,37 %	95,65 %	4,36 %
APG05586	SGN001165	VEGFA	65,59 %	98,58 %	2,03 %
APG05586	SGN001166	VEGFA	10,48 %	94,98 %	5,02 %
APG06015	SGN001322	RelA	N.D.		
APG06015	SGN001325	RelA	N.D.		
APG06015	SGN001327	RelA	N.D.		
APG06015	SGN001333	TRA	N.D.		
APG06015	SGN001334	TRA	0,03	100 %	0 %
APG06015	SGN001335	TRA	N.D.		
APG06015	SGN001351	HAO1	N.D.		
APG06015	SGN001352	HAO1	N.D.		
APG06015	SGN001353	HAO1	N.D.		
APG09344	SGN001321	RelA	0,02	0 %	100 %
APG09344	SGN001324	RelA	N.D.		
APG09344	SGN001329	RelA	N.D.		

ES 3 029 182 T3

(continuación)

NGA	ID de la Guía	Diana génica	Tasa global de edición	Tasa de delección en la muestra	Tasa de inserción en la muestra
APG09344	SGN001336	TRA	N.D.		
APG09344	SGN001337	TRA	N.D.		
APG09344	SGN001338	TRA	N.D.		
APG09344	SGN001354	HAO1	N.D.		
APG09344	SGN001355	HAO1	N.D.		
APG09344	SGN001356	HAO1	N.D.		
APG07991	SGN001312	RelA	2,16	56,61 %	43,39 %
APG07991	SGN001313	RelA	2,64	80,6 %	19,39 %
APG07991	SGN001314	RelA	2,49	32,48 %	67,51 %
APG07991	SGN001339	TRA	6,75	76,47 %	27,42 %
APG07991	SGN001340	TRA	6,66	69,19 %	33,89 %
APG07991	SGN001341	TRA	2,84	60,63 %	39,38 %
APG07991	SGN001357	HAO1	13,34	79,66 %	21,31 %
APG07991	SGN001358	HAO1	0,05	60,78 %	39,22 %
APG07991	SGN001359	HAO1	0,21	80,53 %	19,48 %
APG01868	SGN001315	RelA	10,78	64,07 %	36,36 %
APG01868	SGN001316	RelA	6,26	87,09 %	13,73 %
APG01868	SGN001317	RelA	19,42	62,57 %	38,11 %
APG01868	SGN001342	TRA	4,66	39,32 %	60,68 %
APG01868	SGN001343	TRA	41,39	60,13 %	41,39 %
APG01868	SGN001344	TRA	5,42	60,46 %	39,53 %
APG01868	SGN001360	HAO1	0,38	69,28 %	30,72 %
APG01868	SGN001361	HAO1	9,48	84,92 %	17,07 %
APG01868	SGN001362	HAO1	0,32	28,97 %	71,03 %
APG02998	SGN001318	RelA	1,13	72,78 %	27,22 %
APG02998	SGN001319	RelA	2,55	91,76 %	8,22 %
APG02998	SGN001320	RelA	3,48	43,97 %	56,04 %
APG02998	SGN001345	TRA	2,01	79,92 %	26,06 %
APG02998	SGN001346	TRA	5,87	40,14 %	61,02 %
APG02998	SGN001347	TRA	2,26	59,08 %	40,89 %
APG02998	SGN001363	HAO 1	0,4	16,34 %	83,66 %
APG02998	SGN001364	HAO 1	N.D.		
APG02998	SGN001365	HAO 1	1,14	65,98 %	34,02 %
APG09208	SGN000778	RelA	0,26	0 %	100,00 %
APG09208	SGN000793	AurkB	1,64	50,94 %	49,06 %
APG09208	SGN001213	RelA	0,05	100 %	0,00 %

ES 3 029 182 T3

(continuación)

NGA	ID de la Guía	Diana génica	Tasa global de edición	Tasa de delección en la muestra	Tasa de inserción en la muestra
APG05586	SGN001159	EMX1	40,13	83,86 %	16,13 %
APG05586	SGN001162	TRA	46,12	86,07 %	14,45 %
APG05586	SGN001163	TRA	43,77	93,38 %	7,58 %
APG05586	SGN001164	TRA	8	95,23 %	4,80 %
APG05586	SGN001165	VEGFA	65,59	98,58 %	2,03 %
APG05586	SGN001166	VEGFA	10,48	94,98 %	5,02 %
APG05586	SGN001167	VEGFA	43,8	92,97 %	7,35 %
APG08770	SGN001159	EMX1	11,9	80,98 %	20,13 %
APG08770	SGN001162	TRA	14,57	90,2 %	10,24 %
APG08770	SGN001163	TRA	12,78	95,47 %	5,49 %
APG08770	SGN001164	TRA	3,09	92,84 %	9,00 %
APG08770	SGN001165	VEGFA	23,12	92,75 %	8,00 %
APG08770	SGN001166	VEGFA	10,52	93,45 %	6,92 %
APG08167	SGN001245	VEGFA	N.D.		
APG08167	SGN001246	VEGFA	N.D.		
APG08167	SGN001247	VEGFA	N.D.		
APG08167	SGN001248	RelA	N.D.		
APG08167	SGN001249	RelA	N.D.		
APG08167	SGN001250	RelA	N.D.		
APG08167	SGN001251	AurkB	N.D.		
APG08167	SGN001252	AurkB	N.D.		
APG08167	SGN001253	AurkB	N.D.		
APG01604	SGN001245	VEGFA	69,13	96,82 %	7,23 %
APG01604	SGN001246	VEGFA	4,57	79,07 %	24,78 %
APG01604	SGN001247	VEGFA	18,49	96,17 %	5,09 %
APG01604	SGN001248	RelA	17,04	94,78 %	5,61 %
APG01604	SGN001249	RelA	5,53	87,88 %	14,96 %
APG01604	SGN001250	RelA	21,18	81,7 %	19,19 %
APG01604	SGN001251	AurkB	8,38	84,67 %	15,34 %
APG01604	SGN001252	AurkB	24,74	90,49 %	10,22 %
APG01604	SGN001253	AurkB	0,32	86,44 %	13,56 %
APG03021	SGN001323	RelA	13,73	87,67 %	12,31 %
APG03021	SGN001326	RelA	1,03	78,9 %	21,11 %
APG03021	SGN001328	RelA	7,12	92,18 %	9,68 %
APG03021	SGN001330	TRA	N.D.		
APG03021	SGN001331	TRA	0,45	23,15 %	76,85 %

(continuación)

NGA	ID de la Guía	Diana génica	Tasa global de edición	Tasa de delección en la muestra	Tasa de inserción en la muestra
APG03021	SGN001332	TRA	0,25	49,44 %	50,57 %
APG03021	SGN001348	HAO1	0,08	30,16 %	69,84 %
APG03021	SGN001349	HAO1	0,03	58,33 %	41,67 %
APG03021	SGN001350	HAO1	0,36	70,53 %	29,48 %

Las inserciones y delecciones específicas para las guías respectivas se muestran en las Tablas 6.1-6.3. En estas tablas, la secuencia diana se identifica mediante letras mayúsculas en negrita. Las regiones PAM octaméricas están doblemente subrayadas, con los principales nucleótidos reconocidos en negrita. Las inserciones se identifican con letras minúsculas. Las delecciones se indican con guiones (-). La ubicación de INDELES se calcula a partir del borde proximal del PAM de la secuencia diana, siendo el borde la ubicación 0. La ubicación es positiva (+) si está en el lado diana del borde; la ubicación es negativa (-) si la ubicación está en el lado PAM del borde.

5

Tabla 6.1: Inserciones y deleciones específicas para la Guía 977 usando NGA APG02874

Guía SGN000977 (SEQ ID NO: 157)	N.º de lecturas	% de lecturas	% de indels	Tipo	Ubicación de indels	Tamaño
GGACAGGTTTAAATGGCCCAAGCCTCAGACCCAGG	218651	99,7				
GGACAGGTTTAAATGGCCCAAG-----ACACCCAGG	161	0,07	24,88	Delección	-1	4
GGACAGGTTTAAATGGCCCAAGTAAACCTCAGACCCAGG	153	0,07	23,65	Inserción	+3	3
GGACAGGTTTAAATGGCCCAAG-----G	137	0,06	21,17	Delección	-9	12
GGACAGGTTTAAATGGCCCAAGC-----ACCCAGG	130	0,06	20,09	Delección	-3	5
GGACAGGTTTAAATGGCCCAAGCCTCAGACCCAGG	33	0,02	5,1	Delección	+13	1
GGACAGGTTTAAATGGCCCAAGCCTCAGACCCAGG	33	0,02	5,1	Delección	+5	1

Tabla 6.2: Inserciones y deleciones específicas para la Guía 985 usando NGA APG03850

Guía SGN000985 (SEQ ID NO: 164)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	218367	99,63				
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	268	0,12	33,37	Inserción	3	1
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	204	0,09	25,4	Deleción	4	1
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	140	0,06	17,43	Deleción	-2	7
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	137	0,06	17,06	Deleción	2	1
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	28	0,01	3,49	Deleción	21	1
AGGCTGGGGCCATTAAACCTCTCTCCAGGGGGGGGGTG	26	0,01	3,24	Deleción	-39	88

Tabla 6.3: Inserciones y deleciones específicas para la Guía 793 usando NGA APG09208

Guía SGN000793 (SEQ ID NO: 170)	N.º de lecturas	% de lecturas	% de índices	Tipo	Ubicación de índices	Tamaño
AGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	169578	98,82				
C-----G	471	0,27	23,25	Delección	-17	80
TGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	398	0,23	19,64	Insertión	+3	1
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	190	0,11	9,38	Delección	-15	18
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	133	0,08	6,56	Delección	-10	12
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	110	0,06	5,43	Delección	0	1
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	106	0,06	5,23	Delección	-2	12
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	102	0,06	5,03	Insertión	+3	2
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	92	0,05	4,54	Delección	+1	7
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	82	0,05	4,05	Insertión	0	1
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	61	0,04	3,01	Insertión	+3	38
C-----TCCC	50	0,03	2,47	Delección	-10	44
CTGGAGAGGTTTAAATGGCCCCAGCCTCACAACCGGCTG	48	0,03	2,37	Insertión	+3	29

(continuación)						
Guía SGN000793 (SEQ ID NO: 170)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
CTGGAGAGGTTTTTAATGGCCCCAGCCTCACAaccag cttctggttcgaaactcgagtggaagattggacttc gootgCCCCAGGCTCTGGCCCTCCC	39	0,02	1,92	Insertión	+3	43
CTGGAGAGGTTTTTAATGGCCCCAGCCTC----- -----CC	36	0,02	1,78	Delección	-12	18
CTGGAGAGGTTTTTAATGTCGCCAGCCTCACAgaac tgttcaagtggtgatcacacactgatcacgtga ttgatcatCCCAAGGCTGGCCCTCCC	34	0,02	1,68	Insertión	+3	46
CTGGAGAGGTTTTTAATGGCCCCA----- -----T	27	0,02	1,33	Delección	-23	34
CTGGAGAGGTTTTTAATGGCCCCAGCCTCACAyatg cgaagctgggggtcttatctccacacatatgccca gattoagcaaaaggtatacCCCAGGCTCTGGCCCTCCC	24	0,01	1,18	Insertión	+3	55
CTGGAGAGGTTTTTAATGGCCCCagccctcaacaggtta gctggactatgcatgtgatggctggtgctcaago agccatcttgccctaagaagtgaagagccayyago caaggatagCCCCAGGCTCTGGCCCTCCC	23	0,01	1,14	Insertión	+3	81

Tabla 6.4: Inserciones y deleciones específicas para la Guía 1166 usando NGA APG005586

Guía SGN001166 (SEQ ID NO: 556)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
CGCGCGGACCACGGCTCCTCCGAAGCGAGAACA GCCCAGAAG	197148	89,52				
CGCGCGGACCACGGCTCCTCCGA-----GAACA GCCCAGAAG	3829	1,74	16,59	Deleción	5	0
CGCGCGGACCACGGCTCCTCCGAAG----AACA GCCCAGAAG	1985	0,9	8,6	Deleción	4	-1
CGCGCGGACCACGGCTCCTCCGAA-----CA GCCCAGAAG	1300	0,59	5,63	Deleción	7	-3
CGCGCGGACCACGGCTCCTCCGAA-CGAGAACA GCCCAGAAG	1223	0,56	5,3	Deleción	1	3
CGCGCGGACCACGGCTCCTCCGAAG-AGAACA GCCCAGAAG	1221	0,55	5,29	Deleción	2	1
CGCGCGGACCACGGCTCCTCCGAA----- ----- ----- ----- -----C	936	0,43	4,06	Deleción	117	-113
CGCGCGGACCACGGCTCCTCCGAAGC----- ---CCAGAAG	845	0,38	3,66	Deleción	9	-7
CGCGCGGACCACGGCTCCTCC----- ----- -----T	715	0,32	3,1	Deleción	64	-57
CGCGCGGACCACGGCTCCTCCGAAG----- ----- ----- ----- -----G	615	0,28	2,66	Deleción	140	-137
CGCGCGGACCACGG----- ----- ----- ----- -----T	610	0,28	2,64	Deleción	132	-118
CGCGCGGACCACGGCTCCTCC----CGAGAACA GCCCAGAAG	551	0,25	2,39	Deleción	4	3
CGCGCGGACCACGGCTCCTCCGA-CGAGAACA GCCCAGAAG	497	0,23	2,15	Deleción	2	3
CGCGCGGACCACGGCTCCTCCGAAGTCGAGAACA GCCCAGAAG	438	0,2	1,9	Inserción	1	3
CGCGCGGACCACGGCTCCTCCGAAG-GAGAACA GCCCAGAAG	381	0,17	1,65	Deleción	1	2
G-----gaGGCgg----- -----G	334	0,15	1,45	Deleción	49	-13
CGCGCGGACCACGGCTCCTCCGAA----- -----A	328	0,15	1,42	Deleción	27	-23
CGCGGACCACGGCTCCTCCGAAGagGAGAACA GCCCAGAAG	326	0,15	1,41	Inserción	2	2
CGCGCGGACCACGGCTCCTCCGAAGC----- ----AGAAG	321	0,15	1,39	Deleción	11	-9
CGCGCGGACCACGGCTCCTCCGAAGAG----- -----G	301	0,14	1,3	Deleción	18	-17
CGCGCGGACCACGGCTCCTCCGAAG----- -----AAG	290	0,13	1,26	Deleción	14	-11
CGCGCGGACCACGGCTCC-----GAGAACA GCCCAGAAG	266	0,12	1,15	Deleción	8	2
CGCGCGGACCACGGCTCCTC-----GAGAACA GCCCAGAAG	265	0,12	1,15	Deleción	6	2
CGCGCGGACCACGGCTC-----GAGAACA GCCCAGAAG	232	0,11	1,01	Deleción	9	2
CGCGCGGACCACGGCTCCTCCGAAGcCGAGAACA GCCCAGAAG	217	0,1	0,94	Inserción	1	2
CGCGCGGACCACGGCTCC-----CGAGAACA GCCCAGAAG	199	0,09	0,86	Deleción	7	3
CGCGCGGACCACGGCTCCTCCGAAG----- -----T	192	0,09	0,83	Deleción	17	-14
CGCGCGGACCACGGCTCCTCC----- ---CAGAAG	177	0,08	0,77	Deleción	15	-8

(continuación)

Guía SGN001166 (SEQ ID NO: 556)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
CGCGCGGACCACGGC-----GAGAACAG GCCCAGAAG	158	0,07	0,68	Deleción	11	2
CGCGCGGACCACGGCTCCTCC----- ----- -----T	149	0,07	0,65	Deleción	97	-90
CGCGCGGACCACGGC-----CGAGAACAG GCCCAGAAG	147	0,07	0,64	Deleción	10	3
CGCGCGGACCACGGCTCC----- ---CAGAAG	131	0,06	0,57	Deleción	18	-8
CGCGCGGACCACGGCTCCTCCGAAGC----- ---CAGAAG	127	0,06	0,55	Deleción	10	-8
CGCGCGGACCACGGCTCCTCCGAAG----- ----- -----G	124	0,06	0,54	Deleción	53	-50
CGCGCGGACCACGGCCCCCTCCGA-----GAACA GCCCAGAAG	113	0,05	0,49	Deleción	5	0
CGCGCGGACCACGGCTCCTCC----- ----AGAAG	112	0,05	0,49	Deleción	16	-9
CGCGCGGACCACGGCTCC-----A GCCCAGAAG	104	0,05	0,45	Deleción	14	-4
CGCGCGGA-----GAACA GCCCAGAAG	102	0,05	0,44	Deleción	20	0
CGCGCGGACCACGGC----- --CCAGAAG	101	0,05	0,44	Deleción	20	-7
CGCGCG-----AGAACA GCCCAGAAG	98	0,04	0,42	Deleción	21	1
CGCGCGGACCACGGCTCCTCCG---CGAGAACAG GCCCAGAAG	97	0,04	0,42	Deleción	3	3
CGCGCGGACCACGGCTCCTCCG----- ----- -G	94	0,04	0,41	Deleción	78	-72
CGCGCGGACCACGGCTCCTCCGAAG---GAACA GCCCAGAAG	91	0,04	0,39	Deleción	3	0
CGCGCGGACCACGGCTCCTCCGAAG----- ----AGAAG	90	0,04	0,39	Deleción	12	-9
CGCGC----- --CCAGAAG	80	0,04	0,35	Deleción	30	-7
CGCGCGGACCACGGCTCCTCCGAAGCCAGAACAG GCCCAGAAG	74	0,03	0,32	Inserción	1	1
C-----AGAAG -----AGAAG	73	0,03	0,32	Deleción	56	-9
CGCGCGGACCACGGCTCCTCCGAAG----- ----- -----A	67	0,03	0,29	Deleción	47	-44
CGCGCGGACCACGGCTCCTCCGAA-CGAGAACAG GCCCAGACG	66	0,03	0,29	Deleción	1	3
CGCGCGGACCACGGCTC----- ----AGAAG	63	0,03	0,27	Deleción	20	-9
CGCGCGGACC----- ----CAGAAG	63	0,03	0,27	Deleción	26	-8
CGCGCGGACCACGGCTCCTCCG----- ----- ----- ----- ---T	62	0,03	0,27	Deleción	179	-173
C-----AGAAG -----AGAAG	60	0,03	0,26	Deleción	51	-9
CGCGCGGACCACGGCTCCTCC-----A GCCCAGAAG	59	0,03	0,26	Deleción	11	-4
CGCGCGGACCAC----- ----AGAAG	56	0,03	0,24	Deleción	25	-9
CGCGCGGAC-----GAGAACAG GCCCAGAAG	55	0,02	0,24	Deleción	17	2
CGCGCGGACCACGGCTCCTCT----CGAGAACAG GCCCAGAAG	54	0,02	0,23	Deleción	4	3
CGCGCGGACCACGGCTCCTCCGAAG----- GCCCAGAAG	54	0,02	0,23	Deleción	8	-5

(continuación)

Guía SGN001166 (SEQ ID NO: 556)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
G----- -----AGAACAGCCCAGAAG	53	0,02	0,23	Deleción	39	1
CGCGCGGACCA----- -----GAAG	52	0,02	0,23	Deleción	27	-10
CGCGCGGACACGGCTCC----- -----AGAAG	52	0,02	0,23	Deleción	19	-9
CGCGCGGACACGGCTCCTCCG----GAGAACAA GCCCAGAAG	50	0,02	0,22	Deleción	4	2
CGCGCGGACACGGCTCC----- -CCCAGAAG	48	0,02	0,21	Deleción	16	-6
CGCGCGGACACGGCTCCTCCGA----- -----GAAG	48	0,02	0,21	Deleción	15	-10
CGCGCGGA-----ACA GCCCAGAAG	47	0,02	0,2	Deleción	22	-2
CGCGCGGACACGGCTCCTCCGAAG----ACA GCCCAGAAG	46	0,02	0,2	Deleción	5	-2
CGCGCGGACACGGCTCCT----GCGAGAACAA GCCCAGAAG	45	0,02	0,19	Deleción	5	4
CGCGCGGACC-----CGAGAACAA GCCCAGAAG	45	0,02	0,19	Deleción	15	3
CGCG-----AGAACA GCCCAGAAG	44	0,02	0,19	Deleción	23	1
CGCGCGGACCA-----GAACA GCCCAGAAG	44	0,02	0,19	Deleción	17	0
CG-----A GCCCAGAAG	44	0,02	0,19	Deleción	30	-4
CGCGCGGACACGGCTCCTCCGA----GAACA GCCCAGACG	44	0,02	0,19	Deleción	5	0
CGCGCGGACACGGCTCCTCCGA----GAACA GCCCAGAAG	42	0,02	0,18	Deleción	5	0
CGCGCGGACACGGCTCCTCCTC-CGAGAACAA GCCCAGAAG	42	0,02	0,18	Deleción	2	3
CGCGCGG-----GGAAGAACAA GCCCAGAAG	41	0,02	0,18	Deleción	17	4
CGCGCGGACACGGCTCCTCCGAAG----- -----GAAG	39	0,02	0,17	Deleción	13	-10
CGCGCGGACC-----GAGAACAA GCCCAGAAG	39	0,02	0,17	Deleción	16	2
CGCGCGGACACGGCTCCTCCGAAG----- -----GA-----G	39	0,02	0,17	Deleción	17	-16
GCGGACACGGCTCCTCCGAAGaagCGAGAACAA GCCCAGAAG	38	0,02	0,16	Inserción	3	3
GGACCACGGCTCCTCCGAAGGAGGagGAGAACAA GCCCAGAAG	38	0,02	0,16	Inserción	5	2
CGCGCGGACAC----- -----AAG	38	0,02	0,16	Deleción	27	-11
CGCGCGGACACGGCTCCTCCGAAGC-----A GCCCAGAAG	38	0,02	0,16	Deleción	6	-4
T----- -----T	37	0,02	0,16	Deleción	47	-15
CGCGCGGACACGGCTGTTCTGA----GAACA GCCCAGAAG	37	0,02	0,16	Deleción	5	0
C----- CGAGAACAGCCCAGAAG	36	0,02	0,16	Deleción	31	3
CGCGCGGACACGGC----- -----GAAG	36	0,02	0,16	Deleción	23	-10
CGCGCGGA----- -----AG	35	0,02	0,15	Deleción	32	-12
CGCGCGGACACGG-----AACAA GCCCAGAAG	34	0,02	0,15	Deleción	15	-1
CGCGCGGACACGGCTC-----AACAA GCCCAGAAG	34	0,02	0,15	Deleción	12	-1
CGCGCGGACACGGCCCCCTCC----CGAGAACAA GCCCAGAAG	33	0,01	0,14	Deleción	4	3
G----- -----A	33	0,01	0,14	Deleción	57	-21
CGCGCGGACACGGCTCCTCCGA-GCGAGAACAA GCCCAGAAG	32	0,01	0,14	Deleción	1	4
CGCGCGGACACGTT-----CGAGAACAA GCCCAGAAG	32	0,01	0,14	Deleción	10	3
CGCGA-----AGAACAA GCCCAGAAG	31	0,01	0,13	Deleción	22	1

(continuación)

Guía SGN001166 (SEQ ID NO: 556)	N.º de lecturas	% de lecturas	% de indeles	Tipo	Ubicación de indeles	Tamaño
CGCGCGGACCACGG----- -----AAG	31	0,01	0,13	Delección	25	-11
CGCG-----ACA GCCC-----G	30	0,01	0,13	Delección	36	-20
CGCGCGGACCACGGCTCCTCCGAAG----ACA GCCCAGCAG	30	0,01	0,13	Delección	4	-1
CGCGCGGACCACGGCTCCT-----GAACA GCCCAGAAG	30	0,01	0,13	Delección	9	0
CGCGGACCACGGCTCCTCCGAAGCGGAGAAC GCCCAGAAG	30	0,01	0,13	Inserción	2	1
CG----- AGAACAGCCCAGAAG	29	0,01	0,13	Delección	25	1
CGCGCGGACCACGGCCCCCTCCGAAG---- AACAGCCCAGAAG	29	0,01	0,13	Delección	4	-1
CGCGCGGACCAC----- AGCCCAGAAG	29	0,01	0,13	Delección	20	-4
CGCGCGGACCACGG----- ACAGCCCAGAAG	29	0,01	0,13	Delección	16	-2
CGCGCGGACCACGGCTCCTC----- -----AGAAG	28	0,01	0,12	Delección	17	-9

La robustez de varias nucleasas se comprobó evaluando su capacidad de edición en muchos sitios diana diferentes a través de varios genes. Se probaron más de 40 objetivos en el panel guía ampliado. Todas las proteínas analizadas mostraron una edición robusta en los diversos sitios. Los resultados se muestran en la Tabla 7.

5

Tabla 7: Robustez de algunas nucleasas guiadas por ARN

NGA	Gen	Número de SGN	Tasa global de edición (%)
APG01868	TRA	SGN001684	42,62
		SGN001685	1,43
		SGN001686	8,19
		SGN001687	8,83
		SGN001688	0,21
		SGN001689	0
		SGN001690	0
		SGN001691	20,86
	EMX1	SGN001697	0,93
		SGN001698	0
		SGN001699	18,84
		SGN001700	21,8
		SGN001701	39,94
		SGN001702	2,16
		SGN001703	0,79
		SGN001704	3,49

(continuación)

NGA	Gen	Número de SGN	Tasa global de edición (%)
	EMX1	SGN001705	14,49
		SGN001706	1,04
		SGN001707	2,51
		SGN001708	1,75
		SGN001709	3,62
		SGN001710	0
		SGN001711	13,39
	HAO 1	SGN001713	22,29
		SGN001714	0,67
		SGN001715	0,23
		SGN001716	7,1
		SGN001717	0,68
		SGN001718	0,21
		SGN001719	0
	LDHA	SGN001721	0,17
		SGN001722	1,36
		SGN001723	24,16
		SGN001724	0,2
		SGN001725	4,92
		SGN001726	0
		SGN001727	1,32
		SGN001728	0,28
		SGN001729	1,94
		SGN001730	5,2
		SGN001731	11,28
		SGN001732	6,04
		SGN001733	1,12
		SGN001734	2,11
		SGN001735	7
	TRA	SGN001371	30,52
		SGN001372	26,01
		SGN001373	26,67
		SGN001374	22,94
		SGN001375	21,17
		SGN001376	0
		SGN001377	0

(continuación)

NGA	Gen	Número de SGN	Tasa global de edición (%)
	TRA	SGN001378	30,04
		SGN001379	39,13
		SGN001380	4,93
		SGN001381	17,14
		SGN001382	5,6
APG05586	B2M	SGN001383	36,23
		SGN001384	44,28
		SGN001385	6,33
		SGN001386	5,55
		SGN001387	48,71
		SGN001388	31,78
		SGN001389	41
		SGN001390	48,15
		SGN001391	46,19
		SGN001392	37,22
		SGN001393	31,62
		SGN001394	29,72
	LDHA	SGN001396	32,88
		SGN001397	41,45
		SGN001399	43,28
		SGN001400	2,52
		SGN001401	37,52
		SGN001402	0,37
		SGN001403	53,13
		SGN001404	44,06
		SGN001405	1,46
	HAO 1	SGN001406	21,98
		SGN001407	9,13
		SGN001408	25,06
		SGN001409	43,81
		SGN001410	37,6
		SGN001411	40,75
		SGN001412	18,2
		SGN001413	28,44
		SGN001414	29,39
		SGN001415	0,39

(continuación)

NGA	Gen	Número de SGN	Tasa global de edición (%)
	HAO 1	SGN001416	43,59
	B2M	SGN001592	23,77
		SGN001593	22,7
		SGN001594	23,6
		SGN001595	33,23
		SGN001596	20,88
		SGN001597	1,26
		SGN001598	26,7
		SGN001599	9,41
		SGN001600	28,88
		SGN001601	7,39
		SGN001602	27,8
		SGN001603	6,93
	HAO 1	SGN001616	0
		SGN001617	2,55
		SGN001618	0,6
		SGN001619	6,29
		SGN001620	7,92
		SGN001621	13,03
		SGN001622	5,32
		SGN001623	18,58
		SGN001624	20,03
		SGN001625	1,65
		SGN001626	0,18
		SGN001627	0,73
	LDHA	SGN001640	3,75
		SGN001641	1,13
		SGN001642	14,2
		SGN001643	12,16
		SGN001644	4,9
		SGN001645	4,78
		SGN001646	0,74
		SGN001647	2,89
		SGN001648	0
		SGN001649	0,15
		SGN001650	4,4

(continuación)

NGA	Gen	Número de SGN	Tasa global de edición (%)
	LDHA	SGN001651	7,26
	TRA	SGN001664	0,44
		SGN001665	3,4
		SGN001666	17,24
		SGN001667	2,33
		SGN001668	0,3
		SGN001669	13,68
		SGN001670	0,54
		SGN001671	5,09
		SGN001672	0,22
		SGN001673	12,22
		SGN001674	17,33
		SGN001675	11,69
APG09298	B2M	SGN001383	22,25
		SGN001384	9,5
		SGN001385	2,37
		SGN001386	0,4
		SGN001387	38,03
		SGN001388	17,66
		SGN001389	42,03
		SGN001390	42,88
		SGN001391	16,31
		SGN001392	16,5
		SGN001393	6,27
		SGN001394	17,15
	HAO 1	SGN001406	3,46
		SGN001407	1,87
		SGN001408	10,71
		SGN001409	18,79
		SGN001410	15,09
		SGN001411	14,81
		SGN001412	0,63
		SGN001413	5,92
		SGN001414	5,24
		SGN001415	0,03
		SGN001416	15,41

(continuación)

NGA	Gen	Número de SGN	Tasa global de edición (%)
	LDHA	SGN001395	3,15
		SGN001396	28,53
		SGN001397	11,69
		SGN001399	23,92
		SGN001400	0,89
		SGN001401	33,97
		SGN001402	0
		SGN001403	25,62
		SGN001404	16,96
		SGN001405	0,11
	TRA	SGN001162	44,26
		SGN001163	42,06
		SGN001164	8,1
		SGN001371	21,05
		SGN001372	3,95
		SGN001373	7,8
		SGN001374	9,04
		SGN001375	2,94
		SGN001376	0
		SGN001377	0
		SGN001378	22,14
		SGN001379	27,43
		SGN001380	0,16
		SGN001381	14,18
		SGN001382	2,78

Ejemplo 5: Ingeniería de proteínas de APG05586

- 5 Los restos conservados y variables de APG05586 se identificaron mediante la comparación de APG05586 con homólogos estrechamente relacionados, incluyendo APG08770 (SEQ ID NO:70), APG09882 (establecido como SEQ ID NO: 568 y descrito en la solicitud internacional n.º PCT/US2020/045759, que se incorpora por referencia en su totalidad) y APG01658 (establecido como SEQ ID NO: 569 y descrito en la solicitud internacional n.º PCT/US2020/045759). Se generaron varias variantes que contenían mutaciones en los lugares no conservados para
- 10 identificar los restos críticos en la proteína APG05586. En total, se alteraron 132 restos en diez NGA variantes de ingeniería APG05586 (SEQ ID NO: 570-579). A continuación, se ensayó la actividad de estas diez variantes de APG05586 y de la APG05586 de tipo salvaje en células de mamífero. Utilizando el esqueleto de ARN guía para APG05586 (SEQ ID NO: 66), se comprobó la actividad de los NGA en seis localizaciones genómicas diana (Tabla 8) siguiendo los métodos descritos en el Ejemplo 4. Las secuencias codificantes optimizadas para mamíferos para cada
- 15 variante se proporcionan como SEQ ID NO: 580-589. En la Tabla 8 también se proporcionan secuencias de nucleótidos de cebadores 5' y 3' útiles para la detección de la actividad de edición génica. Las tasas de edición se muestran en la tabla 9.

Tabla 8: Secuencias diana

ID de la Guía	Gen	Secuencia diana (SEQ ID NO:)	ARNgu (SEQ ID NO:)	Cebador en 5' para amplificación	Cebador en 3' para amplificación
SGN001159	EMX1	590	591	592	593
SGN001162	TRA	594	595	596	597
SGN001163	TRA	598	599	596	597
SGN001164	TRA	600	601	596	597
SGN001165	VEGFA	602	603	604	605
SGN001166	VEGFA	606	607	604	605

Tabla 9: Índices de edición para APG05586 y sus variantes

ID de NGA	SEQ ID NO de NGA	SGN001159	SGN001162	SGN001163	SGN001164	SGN001165	SGN001166
APG09298	570	33,40 %	44,26 %	42,06 %	8,10 %	33,94 %	46,19 %
APG06251	571	0 %	43,59 %	48,82 %	5,90 %	25,07 %	35,08 %
APG03066	572	0 %	0 %	0 %	0 %	0 %	0 %
APG01560	573	15,12 %	16,38 %	17,59 %	2,52 %	44,99 %	20,68 %
APG02777	574	0 %	0,08 %	0 %	0 %	0 %	0 %
APG05761	575	24,25 %	21,95 %	21,86 %	10,45 %	23,72 %	20,17 %
APG02479	576	5,61 %	9,25 %	11,00 %	1,45 %	4,53 %	6,34 %
APG08385	577	31,16 %	39,95 %	35,42 %	2,37 %	31,41 %	26,65 %
APG09217	578	29,43 %	41,12 %	44,71 %	4,79 %	38,16 %	31,76 %
APG06657	579	36,81 %	44,20 %	41,54 %	4,44 %	16,29 %	21,86 %
APG05586	63	40,13 %	46,12 %	43,77 %	8,00 %	96,76 %	77,38 %

- 5 La actividad relativa de las variantes sugiere qué lugares de la proteína toleran las mutaciones. La Tabla 10 que se muestra a continuación es un resumen de la actividad de la variante NGA con el número de mutaciones introducidas en comparación con APG005586. "-" es sin actividad; "+" es 1-15 % de edición en al menos 4 de 6 dianas; "++" es 10-25 % de edición en al menos 4 de 6 dianas; "+++" es 20-50 % de edición en al menos 4 de 6 dianas.

10

Tabla 10: Resumen de los cambios en la ingeniería de proteínas e índices de edición

Proteína	Resultado de la actividad	Número de mutaciones
APG09298	+++	10
APG06251	+++	12
APG03066	-	87
APG01560	++	10
APG02777	-	37
APG05761	++	10
APG02479	+	12
APG08385	+++	12
APG09217	+++	14
APG06657	+++	14
APG05586	+++	0

Las variantes APG03066 y APG02777 contenían demasiadas mutaciones para identificar restos específicos importantes para la función, sin embargo, la baja actividad de estas variantes indica que en esta proteína no se toleran cambios extensos en la hélice puente y en el dominio de reconocimiento. Todas las demás variantes contenían 14 o menos mutaciones, lo que permitió identificar restos específicos importantes para la actividad. Basándose en estos resultados, se identificaron varios restos importantes para la función de la proteína. Las mutaciones I305L, V328A, L366I, T368S y V405A provocaron una disminución de la actividad en las variantes ensayadas. Se prevé que todas estas mutaciones se encuentren en el dominio de reconocimiento de la proteína. La disminución de la actividad de APG01560 es el resultado de múltiples cambios que no están localizados en una región específica dentro de la proteína.

Ejemplo 6: Identificación de dianas de enfermedades

Se obtuvo una base de datos de variantes clínicas a partir de la base de datos del NCBI ClinVar, que está disponible a través de Internet en el sitio de Internet del NCBI ClinVar. A partir de esta lista se identificaron polimorfismos de nucleótido único (SNP) patógenos. Usando la información de locus genómicos, se identificaron dianas de CRISPR en la región que solapa y rodea a cada SNP. En la Tabla 11 se enumera una selección de SNP que se pueden corregir usando edición de bases en combinación con las NGA de la invención para dirigirse a la mutación causal ("Mut. CasI"). En la Tabla 11, sólo se enumera un alias de cada enfermedad. El "n.º de RS" corresponde al número de registro del RS a través de la base de datos de SNP en el sitio de Internet del NCBI. El ID de alelo corresponde a un número de registro del alelo causal y el número de registro del cromosoma también proporciona información de referencia de registro que se encuentra en el sitio de Internet del NCBI. La Tabla 11 también proporciona información de la secuencia diana genómica adecuada para la NGA enumerado para cada enfermedad. La información de la secuencia diana también proporciona la secuencia protoespaciadora para la producción del ARN_g necesario para la NGA correspondiente de la invención.

Tabla 11: Enfermedades diana para las NGA

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Simbolo del gen	Diana (SEQ ID NO.)
Enfermedad de Stargardt 1	1800553	APG06622	C>T	22927	NC_000001.10, NC_000001.11	ABCA4	261
Enfermedad de Stargardt 1	1800728	APG06622	A>G	98777	NC_000001.10, NC_000001.11	ABCA4	262
Enfermedad de Stargardt 1	61751374	APG06622	G>A	22933	NC_000001.10, NC_000001.11	ABCA4	263
Enfermedad de Stargardt 1	61750641	APG02874, APG03031, APG09208	G>A	105317	NC_000001.10, NC_000001.11	ABCA4	264
Enfermedad de Stargardt 1	1800553	APG02787	G>A	22927	NC_000001.10, NC_000001.11	ABCA4	265
Enfermedad de Stargardt 1	1800553	APG06007	G>A	22927	NC_000001.10, NC_000001.11	ABCA4	266
Enfermedad de Stargardt 1	1800553	APG03850	G>A	22927	NC_000001.10, NC_000001.11	ABCA4	267
Enfermedad de Stargardt 1	1800553	APG05586	G>A	22927	NC_000001.10, NC_000001.11	ABCA4	268
Enfermedad de Stargardt 1	1800553	APG08167, APG01604	C>T	22927	NC_000001.10, NC_000001.11	ABCA4	269
Hiperinsulinismo familiar	1,51E+08	APG02874, APG03031, APG09208	G>A	24127	NC_000011.9, NC_000011.10	ABCC8	270
Deficiencia de acil-CoA deshidrogenasa de cadena muy larga	1,14E+08	APG06622	T>C	33877	NC_000017.10, NC_000017.11	ACADVL	271
Deficiencia de acil-CoA deshidrogenasa de cadena muy larga	1,14E+08	APG06248	T>C	33877	NC_000017.10, NC_000017.11	ACADVL	272
Deficiencia de acil-CoA deshidrogenasa de cadena muy larga	3,7E+08	APG02874, APG03031, APG09208	G>A	98197	NC_000017.10, NC_000017.11	ACADVL	273
Síndrome de Baraitser-Winter 1	2,82E+08	APG02874, APG03031, APG09208	G>A	38553	NC_000007.13, NC_000007.14	ACTB	274
Inmunodeficiencia grave por deficiencia de ADA	1,22E+08	APG02874, APG03031, APG09208	G>A	16996	NC_000020.10, NC_000020.11	AAF	275
Inmunodeficiencia grave por deficiencia de ADA	1,22E+08	APG02874, APG03031, APG09208	T>C	17004	NC_000020.10, NC_000020.11	AAF	276
Hiperoxaluria primaria	1,22E+08	APG06622	G>A	38436	NC_000002.11, NC_000002.12	AGXT	277

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Cast.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO:)
Trastornos congénitos de la glucosilación	28939378	APG06622	C>T	19763	NC_000016.9,NC_000016.10	ALG1	278
Hipofosfatasa	1.2E+08	APG06622	G>A	28709	NC_000001.10,NC_000001.11	ALPL	279
Hipofosfatasa	1.2E+08	APG02874, APG03031, APG09208	G>A	28709	NC_000001.10,NC_000001.11	ALPL	280
Cáncer colorrectal	1.3E+08	APG06622	C>T	15837	NC_000005.9,NC_000005.10	APC	281
Leucodistrofia metacromática	80338815	APG06622	C>T	18090	NC_000022.10,NC_000022.11	ARSA	282
Enfermedad de Wilson	1.94E+08	APG02874, APG03031, APG09208	G>A	44393	NC_000013.10,NC_000013.11	ATP7B	283
Síndrome cardiolociocutáneo	1.8E+08	APG06622	T>C	29012	NC_000007.13,NC_000007.14	BRAF	284
Síndrome cardiolociocutáneo	1.8E+08	APG06248	T>C	29012	NC_000007.13,NC_000007.14	BRAF	285
Cáncer de mama y/u ovario	41233455	APG06622	G>A	32714	NC_000017.10,NC_000017.11	BRCA1	286
Cáncer de mama y/u ovario	41233465	APG06622	G>A	70268	NC_000017.10,NC_000017.11	BRCA1	287
Cáncer de mama y colorrectal	55770810	APG06622	G>A	70063	NC_000017.10,NC_000017.11	BRCA1	288
Cáncer de mama y/u ovario	62625307	APG06622	G>A	69596	NC_000017.10,NC_000017.11	BRCA1	289
Cáncer de mama y/u ovario	62625308	APG06622	G>A	32710	NC_000017.10,NC_000017.11	BRCA1	290
Cáncer de mama y/u ovario	80356962	APG06622	C>T	70247	NC_000017.10,NC_000017.11	BRCA1	291
Cáncer de mama y/u ovario	41233455	APG02787	C>T	32714	NC_000017.10,NC_000017.11	BRCA1	292
Cáncer de mama y/u ovario	41233455	APG06007	C>T	32714	NC_000017.10,NC_000017.11	BRCA1	293
Cáncer de mama y/u ovario	41233455	APG03850	C>T	32714	NC_000017.10,NC_000017.11	BRCA1	294
Cáncer de mama y/u ovario	41233455	APG05586	C>T	32714	NC_000017.10,NC_000017.11	BRCA1	295
Cáncer de mama y/u ovario	41233455	APG08167, APG01604	G>A	32714	NC_000017.10,NC_000017.11	BRCA1	296
Cáncer de mama y/u ovario	80356962	APG08167, APG01604	C>T	70247	NC_000017.10,NC_000017.11	BRCA1	297
Cáncer de mama y/u ovario	80358163	APG08167, APG01604	T>C	46006	NC_000017.10,NC_000017.11	BRCA1	298

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Cáncer de mama y/u ovario	45580035	APG06622	C>T	67431	NC_000013.10,NC_000013.11	BRCA2	299
Cáncer de mama y/u ovario	80359212	APG06622	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	300
Cáncer de mama y/u ovario	80359003	APG02874, APG03031, APG09208	G>A	67069	NC_000013.10,NC_000013.11	BRCA2	301
Cáncer de mama y/u ovario	80359004	APG02874, APG03031, APG09208	G>A	46672	NC_000013.10,NC_000013.11	BRCA2	302
Cáncer de mama y/u ovario	80359212	APG02787	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	303
Cáncer de mama y/u ovario	80359212	APG06007	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	304
Cáncer de mama y/u ovario	80359212	APG03850	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	305
Cáncer de mama y/u ovario	80359212	APG05586	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	306
Cáncer de mama y/u ovario	80359212	APG08167, APG01604	C>T	67494	NC_000013.10,NC_000013.11	BRCA2	307
Cáncer de mama y/u ovario	80359071	APG08167, APG01604	G>A	67203	NC_000013.10,NC_000013.11	BRCA2	308
Trastornos relacionados con CAPN3	1,21E+08	APG02874, APG03031, APG09208	G>A	32661	NC_000015.9,NC_000015.10	CAPN3;POMT1	309
Deficiencia de CB S	5742905	APG06622	A>G	15159	NC_000021.8,NC_000021.9	CBS	310
Deficiencia de CB S	5742905	APG02874, APG03031, APG09208	T>C	15159	NC_000021.8,NC_000021.9	CBS	311
Deficiencia de CB S	1,22E+08	APG02874, APG03031, APG09208	G>A	15156	NC_000021.8,NC_000021.9	CBS	312
Fibrosis quística	77010898	APG06622	G>A	22168	NC_000007.13,NC_000007.14	CFTR	313
Fibrosis quística	75096551	APG02874, APG03031, APG09208	G>A	33858	NC_000007.13,NC_000007.14	CFTR	314
Fibrosis quística	75527207	APG02787	G>A	22159	NC_000007.13,NC_000007.14	CFTR	315
Fibrosis quística	75527207	APG06007	G>A	22159	NC_000007.13,NC_000007.14	CFTR	316

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Fibrosis quística	75527207	APG03850	G>A	22159	NC_000007.13,NC_000007.14	CFTR	317
Fibrosis quística	75527207	APG05586	G>A	22159	NC_000007.13,NC_000007.14	CFTR	318
Fibrosis quística	78655421	APG02787	G>A	22148	NC_000007.13,NC_000007.14	CFTR	319
Fibrosis quística	78655421	APG06007	G>A	22148	NC_000007.13,NC_000007.14	CFTR	320
Fibrosis quística	78655421	APG03850	G>A	22148	NC_000007.13,NC_000007.14	CFTR	321
Fibrosis quística	78655421	APG05586	G>A	22148	NC_000007.13,NC_000007.14	CFTR	322
Fibrosis quística	75527207	APG08167, APG01604	G>A	22159	NC_000007.13,NC_000007.14	CFTR	323
Miotonía congénita	80356701	APG02874, APG03031, APG09208	T>C	33902	NC_000007.13,NC_000007.14	CLCN1	324
Osteogénesis imperfecta de tipo I	72645321	APG02874, APG03031, APG09208	G>A	414022	NC_000017.10,NC_000017.11	COL1A1	325
Osteogénesis imperfecta de tipo I	72645321	APG02874, APG03031, APG09208	G>A	414022	NC_000017.10,NC_000017.11	COL1A1	326
Síndrome de Alport 1, Recesivo ligado al cromosoma X	1,05E+08	APG02874, APG03031, APG09208	G>A	35796	NC_000023.10,NC_000023.11	COL4A5	327
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG06622	C>T	23992	NC_000001.10,NC_000001.11	CPT2	328
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG06248	C>T	23992	NC_000001.10,NC_000001.11	CPT2	329
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG06007	C>T	23992	NC_000001.10,NC_000001.11	CPT2	330
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG03850	C>T	23992	NC_000001.10,NC_000001.11	CPT2	331
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG05586	C>T	23992	NC_000001.10,NC_000001.11	CPT2	332
Deficiencia de carnitina palmitoiltransferasa II	74315294	APG08167, APG01604	C>T	23992	NC_000001.10,NC_000001.11	CPT2	333
Deficiencia de dopamina beta hidroxilasa	74853476	APG02874, APG03031, APG09208	T>C	16789	NC_000009.11,NC_000009.12	DBH	334
Microcefalia congénita	11555217	APG06622	C>T	34125	NC_000011.9,NC_000011.10	DHCR7	335

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Cast.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Síndrome de Smith-Lemli-Opitz	80338853	APG06622	G>A	21822	NC_000011.9,NC_000011.10	DHCR7	336
Síndrome de Smith-Lemli-Opitz	80338857	APG02874, APG03031, APG09208	G>A	34128	NC_000011.9,NC_000011.10	DHCR7	337
Síndrome de Smith-Lemli-Opitz	11555217	APG06007	G>A	34125	NC_000011.9,NC_000011.10	DHCR7	338
Síndrome de Smith-Lemli-Opitz	11555217	APG03850	G>A	34125	NC_000011.9,NC_000011.10	DHCR7	339
Síndrome de Smith-Lemli-Opitz	11555217	APG05586	G>A	34125	NC_000011.9,NC_000011.10	DHCR7	340
Disautonomía familiar	1,11E+08	APG06248	A>G	21124	NC_000009.11,NC_000009.12	ELP1	341
Hipertirosinemia	80338901	APG06622	G>A	26909	NC_000015.9,NC_000015.10	LEJOS	342
Hipertirosinemia	80338901	APG06248	G>A	26909	NC_000015.9,NC_000015.10	LEJOS	343
Anemia de Fanconi	1,05E+08	APG06622	G>A	27086	NC_000009.11,NC_000009.12	FANCC	344
Síndrome de Marfan	3,98E+08	APG06622	C>T	51454	NC_000015.9,NC_000015.10	FBN1	345
Síndrome de Marfan	7,28E+08	APG06622	A>G	175979	NC_000015.9,NC_000015.10	FBN1	346
Síndrome de Marfan	1,38E+08	APG02874, APG03031, APG09208	G>A	31496	NC_000015.9,NC_000015.10	FBN1	347
Síndrome de Marfan	1,38E+08	APG02874, APG03031, APG09208	G>A	31496	NC_000015.9,NC_000015.10	FBN1	348
Síndrome de Marfan	3,88E+08	APG02874, APG03031, APG09208	G>A	38652	NC_000015.9,NC_000015.10	FBN1	349
Trastornos relacionados con FGFR3	1,22E+08	APG06622	C>T	31371	NC_000004.11,NC_000004.12	FGFR3	350
Enfermedad de almacenamiento de glucógeno de tipo 1A	1801175	APG08167, APG01604	C>T	27037	NC_000017.10,NC_000017.11	G6PC	351
Enfermedad del almacenamiento de glucógeno, tipo II	3,98E+08	APG02874, APG03031, APG09208	G>A	415590	NC_000017.10,NC_000017.11	GAA	352
Deficiencia de UDP-glucosahexosa-1-fosfato uridiltransferasa	75391579	APG06622	A>G	18653	NC_000009.11,NC_000009.12	GALT	353
Aciduria glutárica, tipo 1	1,21E+08	APG02874, APG03031,	G>A	17127	NC_000019.9,NC_000019.10	GCDH	354

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
		APG09208					
Sordera, ligada al X	76434561	APG06622	C>T	53916	NC_000013.10,NC_000013.11	GJB2	355
Sordera, ligada al X	80338945	APG06622	A>G	32055	NC_000013.10,NC_000013.11	GJB2	356
Sordera, ligada al X	80338945	APG02874, APG03031, APG09208	T>C	32055	NC_000013.10,NC_000013.11	GJB2	357
Sordera, ligada al X	80338945	APG02874, APG03031, APG09208	T>C	32055	NC_000013.10,NC_000013.11	GJB2	358
Sordera, ligada al X	1,11E+08	APG02874, APG03031, APG09208	G>A	53902	NC_000013.10,NC_000013.11	GJB2	359
Sordera, ligada al X	1,05E+08	APG06007	G>A	32041	NC_000013.10,NC_000013.11	GJB2	360
Sordera, ligada al X	1,05E+08	APG03850	G>A	32041	NC_000013.10,NC_000013.11	GJB2	361
Sordera, ligada al X	1,05E+08	APG08167, APG01604	C>T	32041	NC_000013.10,NC_000013.11	GJB2	362
Sordera, ligada al X	80338945	APG08167, APG01604	A>G	32055	NC_000013.10,NC_000013.11	GJB2	363
Miopatía por cuerpos de inclusión 2	28937594	APG06622	A>G	21064	NC_000009.11,NC_000009.12	GNE	364
Miopatía por cuerpos de inclusión 2	28937594	APG06248	A>G	21064	NC_000009.11,NC_000009.12	GNE	365
Beta Talasemia	33930165	APG02874, APG03031, APG09208	G>A	30165	NC_000011.9,NC_000011.10	HBB	366
Mucopolisacaridosis Tipo I	1,22E+08	APG06622	G>A	26947	NC_000004.11,NC_000004.12	IDUA	367
Mucopolisacaridosis Tipo I	1,22E+08	APG06622	C>T	26948	NC_000004.11,NC_000004.12	IDUA	368
Mucopolisacaridosis Tipo I	1,22E+08	APG02874, APG03031, APG09208	G>A	26947	NC_000004.11,NC_000004.12	IDUA	369
Síndrome de QT largo congénito	1,99E+08	APG02874, APG03031, APG09208	G>A	67758	NC_000011.9,NC_000011.10	KCNQ1	370
Síndrome de QT largo congénito	1,99E+08	APG02874, APG03031, APG09208	T>C	67776	NC_000011.9,NC_000011.10	KCNQ1	371

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Cast.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO:)
		APG09208					
Hipercolesterolemia familiar	28942080	APG06622	G>A	18735	NC_000019.9;NC_000019.10	LDLR	372
Hipercolesterolemia familiar	1,22E+08	APG06622	C>T	18725	NC_000019.9;NC_000019.10	LDLR	373
Hipercolesterolemia familiar	1,38E+08	APG06622	G>A	171217	NC_000019.9;NC_000019.10	LDLR	374
Hipercolesterolemia familiar	7,46E+08	APG06622	C>T	228192	NC_000019.9;NC_000019.10	LDLR	375
Hipercolesterolemia familiar	7,66E+08	APG06622	G>A	228162	NC_000019.10;NC_000019.9	LDLR	376
Hipercolesterolemia familiar	7,69E+08	APG06622	G>A	228176	NC_000019.10;NC_000019.9	LDLR	377
Hipercolesterolemia familiar	7,46E+08	APG06248	C>T	228192	NC_000019.9;NC_000019.10	LDLR	378
Hipercolesterolemia familiar	7,69E+08	APG06248	G>A	228176	NC_000019.10;NC_000019.9	LDLR	379
Hipercolesterolemia familiar	3,76E+08	APG02874, APG03031, APG09208	G>A	198012	NC_000019.10;NC_000019.9	LDLR	380
Hipercolesterolemia familiar	7,69E+08	APG02874, APG03031, APG09208	G>A	228176	NC_000019.10;NC_000019.9	LDLR	381
Hipercolesterolemia familiar	7,75E+08	APG02874, APG03031, APG09208	T>C	228197	NC_000019.9;NC_000019.10	LDLR	382
Hipercolesterolemia familiar	7,76E+08	APG02874, APG03031, APG09208	G>A	246116	NC_000019.9;NC_000019.10	LDLR	383
Hipercolesterolemia familiar	8,79E+08	APG02874, APG03031, APG09208	T>C	246008	NC_000019.10;NC_000019.9	LDLR	384
Hipercolesterolemia familiar	1,38E+08	APG02787	G>A	171217	NC_000019.9;NC_000019.10	LDLR	385
Hipercolesterolemia familiar	1,38E+08	APG06007	G>A	171217	NC_000019.9;NC_000019.10	LDLR	386
Hipercolesterolemia familiar	1,38E+08	APG03850	G>A	171217	NC_000019.9;NC_000019.10	LDLR	387
Hipercolesterolemia familiar	1,38E+08	APG05586	G>A	171217	NC_000019.9;NC_000019.10	LDLR	388
Hipercolesterolemia familiar	1,38E+08	APG08167, APG01604	G>A	171217	NC_000019.9;NC_000019.10	LDLR	389

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO:)
Hipercolesterolemia familiar	1,22E+08	APG08167, APG01604	C>T	18725	NC_000019.9,NC_000019.10	LDLR	390
Síndrome cardíofaciocutáneo	1,22E+08	APG06622	A>G	28390	NC_000015.9,NC_000015.10	MAP2K1	391
Trastornos relacionados con MECP2	28934906	APG06622	G>A	26850	NC_000023.10,NC_000023.11	MECP2	392
Trastornos relacionados con MECP2	28935468	APG06622	G>A	26863	NC_000023.10,NC_000023.11	MECP2	393
Trastornos relacionados con MECP2	61749721	APG06622	G>A	26868	NC_000023.10,NC_000023.11	MECP2	394
Trastornos relacionados con MECP2	61750240	APG06622	G>A	26854	NC_000023.10,NC_000023.11	MECP2	395
Trastornos relacionados con MECP2	28935468	APG06248	G>A	26863	NC_000023.10,NC_000023.11	MECP2	396
Trastornos relacionados con MECP2	28934906	APG06007	C>T	26850	NC_000023.10,NC_000023.11	MECP2	397
Trastornos relacionados con MECP2	28934906	APG03850	C>T	26850	NC_000023.10,NC_000023.11	MECP2	398
Trastornos relacionados con MECP2	61750240	APG02787	C>T	26854	NC_000023.10,NC_000023.11	MECP2	399
Trastornos relacionados con MECP2	61750240	APG06007	C>T	26854	NC_000023.10,NC_000023.11	MECP2	400
Trastornos relacionados con MECP2	61750240	APG03850	C>T	26854	NC_000023.10,NC_000023.11	MECP2	401
Trastornos relacionados con MECP2	61750240	APG05586	C>T	26854	NC_000023.10,NC_000023.11	MECP2	402
Síndrome de Angelman	61751362	APG02787	C>T	26858	NC_000023.10,NC_000023.11	MECP2	403
Síndrome de Angelman	61751362	APG06007	C>T	26858	NC_000023.10,NC_000023.11	MECP2	404
Síndrome de Angelman	61751362	APG03850	C>T	26858	NC_000023.10,NC_000023.11	MECP2	405
Síndrome de Angelman	61751362	APG05586	C>T	26858	NC_000023.10,NC_000023.11	MECP2	406
Síndrome de Angelman	28934906	APG08167, APG01604	G>A	26850	NC_000023.10,NC_000023.11	MECP2	407
Fiebre mediterránea familiar	28940579	APG06622	A>G	17579	NC_000016.9,NC_000016.10	MEFV	408
Fiebre mediterránea familiar	61752717	APG06622	T>C	17577	NC_000016.9,NC_000016.10	MEFV	409
Fiebre mediterránea familiar	1,05E+08	APG06622	C>T	17588	NC_000016.9,NC_000016.10	MEFV	410
Fiebre mediterránea familiar	61752717	APG06248	T>C	17577	NC_000016.9,NC_000016.10	MEFV	411
Fiebre mediterránea familiar	1,05E+08	APG06248	C>T	17588	NC_000016.9,NC_000016.10	MEFV	412

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casi.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Fiebre mediterránea familiar	1,05E+08	APG02874, APG03031, APG09208	G>A	17588	NC_000016.9,NC_000016.10	MEFV	413
Fiebre mediterránea familiar	28940579	APG08167, APG01604	A>G	17579	NC_000016.9,NC_000016.10	MEFV	414
Enfermedad de Charcot-Marie-Tooth, tipo 2	28940293	APG02874, APG03031, APG09208	T>C	17309	NC_000001.10,NC_000001.11	MFN2	415
Cáncer de colon hereditario no polipósico	63751657	APG02874, APG03031, APG09208	G>A	95331	NC_000003.11,NC_000003.12	MLH1	416
Cáncer de colon hereditario no polipósico	63751711	APG02874, APG03031, APG09208	G>A	95792	NC_000003.11,NC_000003.12	MLH1	417
Acidemia metilmalónica	1,22E+08	APG06622	C>T	16462	NC_000001.10,NC_000001.11	MMACHC	418
Acidemia metilmalónica	1,22E+08	APG06622	C>T	16462	NC_000001.10,NC_000001.11	MMACHC	419
Acidemia metilmalónica	1,22E+08	APG02874, APG03031, APG09208	G>A	16464	NC_000001.10,NC_000001.11	MMACHC	420
Síndrome hereditario de predisposición al cáncer	63750636	APG06622	C>T	96378	NC_000002.11,NC_000002.12	MSH2	421
Síndrome hereditario de predisposición al cáncer	63749843	APG06622	C>T	94826	NC_000002.11,NC_000002.12	MSH6	422
Síndrome hereditario de predisposición al cáncer	7,86E+08	APG06622	C>T	181998	NC_000002.12,NC_000002.11	MSH6	423
Síndrome hereditario de predisposición al cáncer	63750741	APG02874, APG03031, APG09208	T>C	94663	NC_000002.11,NC_000002.12	MSH6	424
Síndrome hereditario de predisposición al cáncer	63749843	APG02787	C>T	94826	NC_000002.11,NC_000002.12	MSH6	425
Síndrome hereditario de predisposición al cáncer	63749843	APG06007	C>T	94826	NC_000002.11,NC_000002.12	MSH6	426
Síndrome hereditario de predisposición al cáncer	63749843	APG03850	C>T	94826	NC_000002.11,NC_000002.12	MSH6	427
Síndrome hereditario de predisposición al cáncer	63749843	APG05586	C>T	94826	NC_000002.11,NC_000002.12	MSH6	428
Síndrome hereditario de predisposición al cáncer	63751017	APG02787	C>T	94786	NC_000002.11,NC_000002.12	MSH6	429

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Síndrome hereditario de predisposición al cáncer	63751017	APG06007	C>T	94786	NC_000002.11,NC_000002.12	MSH6	430
Síndrome hereditario de predisposición al cáncer	63751017	APG03850	C>T	94786	NC_000002.11,NC_000002.12	MSH6	431
Síndrome hereditario de predisposición al cáncer	63751017	APG05586	C>T	94786	NC_000002.11,NC_000002.12	MSH6	432
Síndrome hereditario de predisposición al cáncer	63751017	APG08167, APG01604	C>T	94786	NC_000002.11,NC_000002.12	MSH6	433
Poliposis asociada a MUTYH	34612342	APG06622	T>C	20332	NC_000001.10,NC_000001.11	MUTYH	434
Poliposis asociada a MUTYH	36053993	APG06622	C>T	20333	NC_000001.10,NC_000001.11	MUTYH	435
Poliposis asociada a MUTYH	36053993	APG02787	G>A	20333	NC_000001.10,NC_000001.11	MUTYH	436
Poliposis asociada a MUTYH	36053993	APG06007	G>A	20333	NC_000001.10,NC_000001.11	MUTYH	437
Poliposis asociada a MUTYH	36053993	APG03850	G>A	20333	NC_000001.10,NC_000001.11	MUTYH	438
Poliposis asociada a MUTYH	36053993	APG05586	G>A	20333	NC_000001.10,NC_000001.11	MUTYH	439
Poliposis asociada a MUTYH	36053993	APG08167, APG01604	C>T	20333	NC_000001.10,NC_000001.11	MUTYH	440
Poliposis asociada a MUTYH	34612342	APG08167, APG01604	T>C	20332	NC_000001.10,NC_000001.11	MUTYH	441
Hiperinmunoglobulina D con fiebre periódica	28934897	APG06622	G>A	26968	NC_000012.11,NC_000012.12	MVK	442
Trastornos relacionados con MYBPC3	2E+08	APG06622	C>T	174776	NC_000011.9,NC_000011.10	MYBPC3	443
Trastornos relacionados con MYBPC3	3.88E+08	APG06622	G>A	45725	NC_000011.9,NC_000011.10	MYBPC3	444
Trastornos relacionados con MYBPC3	3.98E+08	APG06622	C>T	51962	NC_000011.9,NC_000011.10	MYBPC3	445
Trastornos relacionados con MYBPC3	2E+08	APG06248	C>T	174776	NC_000011.9,NC_000011.10	MYBPC3	446
Trastornos relacionados con MYBPC3	1.88E+08	APG02874, APG03031, APG09208	T>C	45267	NC_000011.9,NC_000011.10	MYBPC3	447
Trastornos relacionados con MYBPC3	2E+08	APG02874, APG03031, APG09208	G>A	174776	NC_000011.9,NC_000011.10	MYBPC3	448
Trastornos relacionados con MYBPC3	3.98E+08	APG02874, APG03031, APG09208	G>A	51820	NC_000011.9,NC_000011.10	MYBPC3	449

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Cast.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Trastornos relacionados con MYBPC3	3.98E+08	APG02874, APG03031, APG09208	G>A	51962	NC_000011.9,NC_000011.10	MYBPC3	450
Miocardopatía	3218716	APG06622	C>T	52071	NC_000014.8,NC_000014.9	MYH7	451
Miocardopatía	36211715	APG06622	C>T	29159	NC_000014.8,NC_000014.9	MYH7	452
Miocardopatía	1.22E+08	APG06622	G>A	29128	NC_000014.8,NC_000014.9	MYH7	453
Miocardopatía	3.72E+08	APG06622	C>T	52045	NC_000014.8,NC_000014.9	MYH7	454
Miocardopatía	3.98E+08	APG02874, APG03031, APG09208	T>C	52276	NC_000014.8,NC_000014.9	MYH7	455
Miocardopatía	3.72E+08	APG06607	G>A	52045	NC_000014.8,NC_000014.9	MYH7	456
Miocardopatía	3.72E+08	APG03850	G>A	52045	NC_000014.8,NC_000014.9	MYH7	457
Sordera, autosómico recesivo	1,11E+08	APG02874, APG03031, APG09208	G>A	52388	NC_000011.9,NC_000011.10	MYO7A	458
Enfermedades genéticas congénitas	80358259	APG06622	A>G	18006	NC_000018.9,NC_000018.10	NPC1	459
Enfermedades genéticas congénitas	80358259	APG02874, APG03031, APG09208	T>C	18006	NC_000018.9,NC_000018.10	NPC1	460
Enfermedades genéticas congénitas	1.2E+08	APG02874, APG03031, APG09208	G>A	18010	NC_000018.9,NC_000018.10	NPC1	461
Fenilcetonuria	5030851	APG06622	G>A	15628	NC_000012.11,NC_000012.12	PAH	462
Fenilcetonuria	5030858	APG06622	G>A	15616	NC_000012.11,NC_000012.12	PAH	463
Hiperfenilalaninemia	5030860	APG06622	T>C	15632	NC_000012.11,NC_000012.12	PAH	464
Fenilcetonuria	62516101	APG02874, APG03031, APG09208	G>A	15658	NC_000012.11,NC_000012.12	PAH	465
Hiperfenilalaninemia, no-pku	62844499	APG02874, APG03031, APG09208	G>A	15656	NC_000012.11,NC_000012.12	PAH	466
Fenilcetonuria	5030858	APG02787	C>T	15616	NC_000012.11,NC_000012.12	PAH	467

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casi.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Fenilcetonuria	5030858	APG06007	C>T	15616	NC_000012.11,NC_000012.12	PAH	468
Fenilcetonuria	5030858	APG03850	C>T	15616	NC_000012.11,NC_000012.12	PAH	469
Fenilcetonuria	5030858	APG05586	C>T	15616	NC_000012.11,NC_000012.12	PAH	470
Hiperfenilalaninemia, no-pku	5030860	APG08167, APG01604	T>C	15632	NC_000012.11,NC_000012.12	PAH	471
Hiperfenilalaninemia, no-pku	62642937	APG08167, APG01604	G>A	15667	NC_000012.11,NC_000012.12	PAH	472
Cáncer familiar de mama	1,8E+08	APG06622	G>A	132139	NC_000016.10,NC_000016.9	PALB2	473
Cáncer familiar de mama	1,8E+08	APG02874, APG03031, APG09208	G>A	132185	NC_000016.10,NC_000016.9	PALB2	474
Trastorno de la biogénesis peroxisomal 1B	61750420	APG08167, APG01604	C>T	22555	NC_000007.13,NC_000007.14	PEX1	475
Inmunodeficiencia 14	3,98E+08	APG06622	G>A	94255	NC_000001.10,NC_000001.11	PIK3CD	476
Displasia renal poliquística	1,38E+08	APG06622	G>A	19147	NC_000006.11,NC_000006.12	PKHD1	477
Síndrome de glucoproteína deficiente en carbohidratos tipo I	28936415	APG02787	G>A	22745	NC_000016.9,NC_000016.10	PMM2	478
Síndrome de glucoproteína deficiente en carbohidratos tipo I	28936415	APG06007	G>A	22745	NC_000016.9,NC_000016.10	PMM2	479
Síndrome de glucoproteína deficiente en carbohidratos tipo I	28936415	APG03850	G>A	22745	NC_000016.9,NC_000016.10	PMM2	480
Síndrome de glucoproteína deficiente en carbohidratos tipo I	28936415	APG05586	G>A	22745	NC_000016.9,NC_000016.10	PMM2	481
Síndrome de glucoproteína deficiente en carbohidratos tipo I	28936415	APG08167, APG01604	G>A	22745	NC_000016.9,NC_000016.10	PMM2	482
Afección relacionada con POLG	1,14E+08	APG06622	C>T	28535	NC_000015.9,NC_000015.10	POLG	483
Afección relacionada con POLG	1,14E+08	APG06622	C>T	28541	NC_000015.9,NC_000015.10	POLG	484
Afección relacionada con POLG	1,14E+08	APG02874, APG03031, APG09208	G>A	28535	NC_000015.9,NC_000015.10	POLG	485
Afección relacionada con POLG	1,14E+08	APG06007	G>A	28535	NC_000015.9,NC_000015.10	POLG	486
Afección relacionada con POLG	1,14E+08	APG03850	G>A	28535	NC_000015.9,NC_000015.10	POLG	487

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO:)
Afección relacionada con POLG	1,14E+08	APG08167, APG01604	C>T	28541	NC_000015.9,NC_000015.10	POLG	488
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG06622	G>A	23943	NC_000001.10,NC_000001.11	PPT1	489
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG02787	C>T	23943	NC_000001.10,NC_000001.11	PPT1	490
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG06007	C>T	23943	NC_000001.10,NC_000001.11	PPT1	491
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG03850	C>T	23943	NC_000001.10,NC_000001.11	PPT1	492
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG05586	C>T	23943	NC_000001.10,NC_000001.11	PPT1	493
Lipofuscinosis neuronal cerioidea 1	1,38E+08	APG08167, APG01604	G>A	23943	NC_000001.10,NC_000001.11	PPT1	494
Miocardopatía	1,22E+08	APG06622	C>T	21885	NC_000007.13,NC_000007.14	PRKAG2	495
Síndrome de Cowden	1,22E+08	APG06622	C>T	22852	NC_000010.10,NC_000010.11	PTEN	496
Trastorno relacionado con PTPN11	1,22E+08	APG06622	C>T	28370	NC_000012.11,NC_000012.12	PTPN 11	497
Linfoma de leucemia linfoblástica B, sin subtipo ICD-O	1,22E+08	APG06622	A>G	28372	NC_000012.11,NC_000012.12	PTPN 11	498
Trastorno relacionado con PTPN11	3,98E+08	APG06622	A>G	49032	NC_000012.11,NC_000012.12	PTPN 11	499
Trastorno relacionado con PTPN11	1,22E+08	APG06248	C>T	28370	NC_000012.11,NC_000012.12	PTPN 11	500
Trastorno relacionado con PTPN11	1,22E+08	APG06248	A>G	28379	NC_000012.11,NC_000012.12	PTPN11	501
Trastorno relacionado con PTPN 11	28933386	APG02787	A>G	28365	NC_000012.11,NC_000012.12	PTPN11	502
Trastorno relacionado con PTPN11	28933386	APG06007	A>G	28365	NC_000012.11,NC_000012.12	PTPN11	503
Trastorno relacionado con PTPN11	28933386	APG03850	A>G	28365	NC_000012.11,NC_000012.12	PTPN11	504
Trastorno relacionado con PTPN11	28933386	APG05586	A>G	28365	NC_000012.11,NC_000012.12	PTPN11	505
Trastorno relacionado con PTPN11	1,22E+08	APG02787	A>G	28372	NC_000012.11,NC_000012.12	PTPN11	506
Trastorno relacionado con PTPN11	1,22E+08	APG06007	A>G	28372	NC_000012.11,NC_000012.12	PTPN11	507
Trastorno relacionado con PTPN11	1,22E+08	APG03850	A>G	28372	NC_000012.11,NC_000012.12	PTPN11	508
Trastorno relacionado con PTPN11	1,22E+08	APG05586	A>G	28372	NC_000012.11,NC_000012.12	PTPN11	509
Trastorno relacionado con PTPN11	28933386	APG08167, APG01604	A>G	28365	NC_000012.11,NC_000012.12	PTPN11	510

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO:)
Enfermedad del almacenamiento de glucógeno	1.17E+08	APG06622	G>A	17337	NC_000011.9,NC_000011.10	PYGM	511
Enfermedad del almacenamiento de glucógeno	1.17E+08	APG06248	G>A	17337	NC_000011.9,NC_000011.10	PYGM	512
Cáncer de mama-ovario, familiar 4	3.88E+08	APG06622	G>A	39241	NC_000017.10,NC_000017.11	RAD51D	513
Miocardíopatía dilatada 1DD	2.68E+08	APG02874, APG03031, APG09208	G>A	15310	NC_000010.10,NC_000010.11	RBM20	514
Trastornos relacionados con RET	74799832	APG06622	T>C	28958	NC_000010.10,NC_000010.11	RET	515
Trastornos relacionados con RET	74799832	APG02874, APG03031, APG09208	T>C	28958	NC_000010.10,NC_000010.11	RET	516
Trastornos relacionados con RYR1	1.18E+08	APG06622	C>T	28003	NC_000019.9,NC_000019.10	RYR1	517
Trastornos relacionados con RYR1	2.01E+08	APG06622	C>T	169564	NC_000019.9,NC_000019.10	RYR1	518
Trastornos relacionados con RYR1	1.18E+08	APG02874, APG03031, APG09208	G>A	76888	NC_000019.9,NC_000019.10	RYR1	519
Trastornos relacionados con RYR1	1.18E+08	APG02874, APG03031, APG09208	G>A	76888	NC_000019.9,NC_000019.10	RYR1	520
Trastornos relacionados con RYR1	1.18E+08	APG02874, APG03031, APG09208	G>A	76835	NC_000019.9,NC_000019.10	RYR1	521
Trastornos relacionados con RYR1	1.18E+08	APG02874, APG03031, APG09208	T>C	28014	NC_000019.9,NC_000019.10	RYR1	522
Síndrome de Shwachman	1.14E+08	APG06622	A>G	18235	NC_000007.13,NC_000007.14	SBDS	523
Síndrome de Brugada	1.38E+08	APG06622	C>T	24416	NC_000003.11,NC_000003.12	SCN5A	524
Síndrome de Brugada	28937316	APG02874, APG03031, APG09208	G>A	24408	NC_000003.11,NC_000003.12	SCN5A	525
Síndrome de Brugada	45546039	APG02874, APG03031, APG09208	G>A	48043	NC_000003.11,NC_000003.12	SCN5A	526
Síndrome de Brugada	72549410	APG02874, APG03031,	G>A	78547	NC_000003.11,NC_000003.12	SCN5A	527

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
		APG09208					
Síndrome de Cowden 3	80338844	APG06622	C>T	21935	NC_000011.9,NC_000011.10	SDHD	528
Síndrome de Cowden 3	80338844	APG06248	C>T	21935	NC_000011.9,NC_000011.10	SDHD	529
Deficiencia de alfa-1-antitripsina	28929474	APG06622	C>T	33006	NC_000014.8,NC_000014.9	SERPINA1	530
Deficiencia de alfa-1-antitripsina	28929474	APG02787	G>A	33006	NC_000014.8,NC_000014.9	SERPINA1	531
Deficiencia de alfa-1-antitripsina	28929474	APG03850	G>A	33006	NC_000014.8,NC_000014.9	SERPINA1	532
Deficiencia de alfa-1-antitripsina	28929474	APG05586	G>A	33006	NC_000014.8,NC_000014.9	SERPINA1	533
Deficiencia de alfa-1-antitripsina	28929474	APG08167, APG01604	C>T	33006	NC_000014.8,NC_000014.9	SERPINA1	534
Distrofia muscular de la cintura y extremidades, tipo 2D	28933693	APG06622	C>T	24476	NC_000017.10,NC_000017.11	SGCA	535
Mucopolisacaridosis, MPS-III-A	1,05E+08	APG06622	C>T	20146	NC_000017.10,NC_000017.11	SGSH	536
Mucopolisacaridosis, MPS-III-A	1,05E+08	APG02874, APG03031, APG09208	G>A	20146	NC_000017.10,NC_000017.11	SGSH	537
Síndrome de Noonan	2,68E+08	APG02787	A>G	21860	NC_000010.10,NC_000010.11	SHOC2	538
Síndrome de Noonan	2,68E+08	APG06007	A>G	21860	NC_000010.10,NC_000010.11	SHOC2	539
Síndrome de Noonan	2,68E+08	APG03850	A>G	21860	NC_000010.10,NC_000010.11	SHOC2	540
Síndrome de Noonan	2,68E+08	APG05586	A>G	21860	NC_000010.10,NC_000010.11	SHOC2	541
Síndrome de Noonan	2,68E+08	APG08167, APG01604	A>G	21860	NC_000010.10,NC_000010.11	SHOC2	542
Trastornos relacionados con SLC26A2	1,05E+08	APG06622	C>T	19128	NC_000005.9,NC_000005.10	SLC26A2	543
Trastornos relacionados con SLC26A2	1,11E+08	APG02874, APG03031, APG09208	G>A	52666	NC_000007.13,NC_000007.14	SLC26A4	544
Miocardopatía hipertrófica familiar 2	7,28E+08	APG02874, APG03031, APG09208	G>A	172354	NC_000001.10,NC_000001.11	TNNT2	545
Enfermedad de Charcot-Marie-Tooth de tipo 2C	3,98E+08	APG02874, APG03031, APG09208	G>A	48018	NC_000012.11,NC_000012.12	TRPV4	546

(continuación)

Enfermedad	n.º de RS	NGA	Mut. Casl.	ID del Alelo	Registro del Cromosoma	Símbolo del gen	Diana (SEQ ID NO.)
Displasia cortical focal de tipo II	28934872	APG06622	G>A	27436	NC_000016.9,NC_000016.10	TSC2	547
Amiloidosis por transtiretina amiloidogénica	76992529	APG06622	G>A	28465	NC_000018.9,NC_000018.10	TTR	548
Amiloidosis por transtiretina amiloidogénica	76992529	APG06248	G>A	28465	NC_000018.9,NC_000018.10	TTR	549
Albinismo oculocutáneo	1,05E+08	APG06622	C>T	18816	NC_000011.9,NC_000011.10	TYR	550
Albinismo oculocutáneo	1,22E+08	APG02874, APG03031, APG09208	G>A	18814	NC_000011.9,NC_000011.10	TYR	551
Enfermedad de Von Willebrand	41276738	APG06622	C>T	15335	NC_000012.11,NC_000012.12	VWF	552

Ejemplo 7: Direccionamiento a mutaciones responsables del síndrome de Hurler

A continuación se describe un posible tratamiento para el síndrome de Hurler, también denominado MPS-1, utilizando un sistema de edición de bases dirigido por ARN que corrige una mutación responsable del síndrome de Hurler en una gran proporción de pacientes con la enfermedad. Este enfoque utiliza una proteína de fusión de edición de bases guiada por ARN y que puede empaquetarse en un único vector de AAV para la administración a una amplia gama de tipos de tejido. Dependiendo de los elementos regulatorios exactos y del dominio editor de base utilizado, también puede ser posible diseñar técnicamente un único vector que codifique tanto la proteína de fusión de edición de bases como un ARN de guía único para dirigirse al locus de la enfermedad.

Ejemplo 7.1; Identificando de NGA con un PAM ideal

La enfermedad genética MPS-1 es una enfermedad de almacenamiento lisosómico caracterizada a nivel molecular por la acumulación de dermatán sulfato y heparán sulfato en los lisosomas. La enfermedad es generalmente un trastorno genético hereditario provocado por mutaciones en el gen IDUA (secuencia de referencia del NCBI NG_008103. 1), que codifica la α -L-iduronidasa. La enfermedad es el resultado de una deficiencia de α -L-iduronidasa. Las mutaciones de IDUA más comunes encontradas en estudios de individuos originarios del norte de Europa son W402X y Q70X, mutaciones terminadoras que dan como resultado en la terminación prematura de la traducción (Bunge *et al.* (1994), Hum. Mol. Genet, 3(6): 861-866, incorporada en el presente documento por referencia). La inversión de un único nucleótido restablecería la secuencia codificante de tipo silvestre y daría como resultado la expresión de proteínas controlada por los mecanismos reguladores endógenos del locus genético.

La mutación W402X del gen humano *Idua* representa una alta proporción de casos de MPS-1H. Los editores de bases pueden dirigirse a una ventana de secuencia estrecha con respecto al sitio de unión del componente de protoespaciador del ARN de guía y, por tanto, la presencia de una secuencia PAM a una distancia específica del locus diana es esencial para el éxito de la estrategia. Dadas las limitaciones de que la mutación diana debe estar en la cadena no diana (CND) expuesta durante la interacción de la proteína de edición de bases y que la huella del dominio de NGA bloqueará el acceso a la región cercana al PAM, se cree que un locus accesible está a 10-30 pb del PAM. Para evitar la edición y mutagénesis de otras bases de adenosina cercanas en esta ventana, se criban diferentes enlazadores. La ventana ideal está a 12-16 pb del PAM.

Una secuencia PAM compatible con APG02874, APG09208 y APG05586 es fácilmente evidente en el locus genético. Estas nucleasas tienen una secuencia PAM de 5'-nnnnCC-3' (SEQ ID NO: 35), 5'-nnnnC-3' (SEQ ID NO: 62) y 5'-nnRYA-3' (SEQ ID NO: 69), respectivamente, y son de tamaño compacto, lo que posiblemente permite su administración a través de un único vector AAV. Este enfoque de administración otorga múltiples ventajas en relación con otros, tales como el acceso a una amplia gama de tejidos (hígado, músculo, SNC) y un perfil de seguridad y técnicas de fabricación bien establecidos.

Cas9 de *S. pyogenes* (SpyCas9) precisa una secuencia de PAM de NGG (SEQ ID NO: 256), que está presente cerca del locus W402X, pero el tamaño de SpyCas9 impide el empaquetamiento en un único vector de AAV y, por tanto, se pierden las ventajas antes mencionadas de este enfoque. Si bien se puede emplear una estrategia de administración doble (por ejemplo, Ryu *et al.* (2018), Nat. Biotechnol., 36(6): 536-539, incorporado en el presente documento por referencia), se añadiría una complejidad y un costo de fabricación significativos. Además, la administración de un vector vírico doble disminuye significativamente la eficacia de la corrección génica, ya que una edición satisfactoria en una célula determinada precisa la infección con ambos vectores y el ensamblaje de la proteína de fusión en la célula.

Un ortólogo de Cas9 de *S. aureus* (SauCas9) usado habitualmente es de tamaño considerablemente más pequeño en relación con SpyCas9 pero tiene un requisito PAM más complejo - NGRRT (SEQ ID NO: 257). Esta secuencia no está dentro de un intervalo que se espera que sea útil para la edición de bases del locus causante.

Ejemplo 7.2: Construcciones de fusión de NGA y secuencias de ARNg

Se produce una secuencia de ADN que codifica una proteína de fusión con los siguientes dominios utilizando técnicas convencionales de biología molecular: 1) un dominio de NGA con mutaciones que inactivan la actividad de escisión de ADN ("sin actividad" o "nickasa"); 2) una adenosina desaminasa útil para la edición de bases. Todas las construcciones descritas en la siguiente tabla comprenden una proteína de fusión con el dominio activo de edición de bases, en este ejemplo, ADAT (SEQ ID NO: 211), fusionada operativamente al extremo N-terminal de la NGA sin actividad APG02874 (SEQ ID NO: 214), APG09208 (SEQ ID NO: 216) y APG005586 (SEQ ID NO: 567). También pueden utilizarse otras adenosina desaminasas útiles para la edición de bases de ADN (véase, por ejemplo, la solicitud PCT PCT/US2019/068079). Se sabe en la técnica que también se podría fabricar una proteína de fusión con la enzima de edición de bases en el extremo carboxiterminal de la NGA. Además, la NGA y el editor de bases de la proteína de fusión normalmente están separados por una secuencia de aminoácidos enlazadora. Se sabe en la técnica que las longitudes de los enlazadores convencionales oscilan de 15-30 aminoácidos. Además, se sabe en la técnica que determinadas proteínas de fusión entre una NGA y una enzima de edición de bases también pueden comprender al menos un dominio inhibidor de uracilo glucosilasa (IUG) (SEQ ID NO: 212), que puede aumentar la eficacia de edición de bases (Patente de EE. UU. N.º 10.167.457, incorporada en el presente documento por referencia). Por lo tanto,

una proteína de fusión puede comprender NGA APG02874, APG09208 o una variante de la misma, una adenosina desaminasa y, opcionalmente, al menos un IUG.

Tabla 12: Construcción para la edición de bases dirigida a ARN

SEQ ID NO.	Construcción	NGA	Sin actividad (D) o nickasa (N)	Editor básico
213	Nuc-ADAT-Enlazador-d APG02874 - Enlazador-SV40	APG02874	D	ADAT
215	Nuc-ADAT-Enlazador-d APG09208 - Enlazador-SV40	APG09208	D	ADAT
565	Nuc-ADAT-Enlazador-d APG05586 - Enlazador-SV40	APG05586	D	ADAT

Los sitios de edición accesibles de una NGA están determinados por la secuencia PAM. Al combinar una NGA con un dominio de edición de bases, el resto diana para la edición debe residir en la cadena no diana (CND), ya que la CND es monocatenaria mientras que la NGA está asociada al locus. La evaluación de una serie de nucleasas y los ARN de guía correspondientes facilita seleccionar la herramienta de edición génica más adecuada para este locus en particular. Varias posibles secuencias PAM a las que pueden dirigirse las construcciones descritas anteriormente en el gen *Idua* humano se encuentran en la proximidad del nucleótido mutante responsable de la mutación W402X. Una secuencia que codifica un transcrito de ARN guía que contiene 1) un "espaciador" que es complementario a la cadena de ADN no codificante en el locus de la enfermedad; y 2) también se produce la secuencia de ARN necesaria para la asociación del ARN guía con la NGA. Dicho ARN puede estar codificado por, por ejemplo, SEQ ID NO: 217 para el sistema APG02874 NGA, SEQ ID NO: 218 para el sistema APG09208 NGA o SEQ ID NO: 566 para el sistema APG05586 NGA. Estas moléculas de ARN o ARN similar, que pueden ser diseñadas un experto en la materia, se pueden evaluar para determinar su eficacia para dirigir los editores de bases anteriores al locus de interés.

Ejemplo 7. 3: Ensayo de actividad en células de pacientes con enfermedad de Hurler

Para verificar la estrategia de genotipo y evaluar las construcciones descritas anteriormente, se utilizan fibroblastos de pacientes con enfermedad de Hurler. Se diseña un vector que contiene promotores apropiados secuencia arriba de la secuencia codificante de la proteína de fusión y la secuencia codificante del ARN para la expresión de éstas en células humanas, de manera similar a los vectores descritos en el Ejemplo 4. Se reconoce que también se pueden usar promotores y otros elementos de ADN (por ejemplo, potenciadores o terminadores) que son conocidos por sus altos niveles de expresión en células humanas o que pueden expresarse específicamente bien en fibroblastos. El vector se transfecta en los fibroblastos usando técnicas convencionales, por ejemplo, una transfección similar a lo que se describe en el Ejemplo 4. Como alternativa, se puede utilizar electroporación. Las células se cultivan durante 1-3 días. El ADN genómico (ADNg) se aísla mediante técnicas convencionales. La eficacia de edición se determina realizando un ensayo de genotipado por qPCR y/o secuenciación de próxima generación sobre el ADNg purificado, como se describe adicionalmente a continuación.

El análisis por qPCR Taqman™ utiliza sondas específicas para el alelo mutante y de tipo silvestre. Estas sondas portan fluoróforos que se resuelven mediante sus propiedades de excitación y/o emisión espectral utilizando un instrumento de qPCR. Se puede obtener en el mercado un kit de genotipado que contiene cebadores y sondas de PCR (es decir, ensayo de genotipado de SNP Thermo Fisher Taqman™ ID C 27862753_10 para SNP ID rs121965019) o diseñarlo. En la Tabla 13 se muestra un ejemplo de un conjunto de cebador y sonda diseñado.

Tabla 13: Cebadores y sondas de RT-PCR

Descripción	Secuencia	SEQ ID NO.
Cebador de amplificación directo	5'-GACTCCTTCACCAAG-3'	219
Cebador de amplificación inverso	5'-GTAGATCAGCACCG-3'	220
Sonda de tipo silvestre	5'-CTCTGGGCCGAAGT-3'	221
Sonda de W402X	5'-CTCTAGGCCGAAGT-3'	222

Después del experimento de edición, el ADNg se somete a análisis de qPCR utilizando métodos convencionales y los cebadores y sondas descritos anteriormente. Los resultados se muestran en la Tabla 14. Este sistema *in vitro* se puede utilizar para evaluar construcciones de manera conveniente y elegir una con alta eficacia de edición para estudios posteriores. Los sistemas se evaluarán en comparación con células con y sin la mutación W402X, y preferentemente con algunas que sean heterocigóticas para esta mutación. Los valores de Ct se compararán con un gen de referencia o con la amplificación total del locus utilizando un colorante tal como Sybr green.

Tabla 14: Resultados esperados de qPCR

Genotipo	Transfectado con editor de bases	Resultado de esperados de PCR
Idua ^{TS/TS}	No	TS homocigótico
Idua ^{TS/W402X}	No	Heterocigótico: 50 % TS, 50 % de W402X
Idua ^{W402X/W402X}	No	W402X homocigótico
Idua ^{W402X/W402X}	Sí	Variable

Los tejidos también se pueden analizar mediante secuenciación de próxima generación. Se pueden usar sitios de unión de cebadores, tales como los que se muestran a continuación (Tabla 15), u otros sitios de unión de cebadores adecuados que pueden ser identificados por un experto en la técnica. Después de la amplificación por PCR, los productos conteniendo secuencias protuberantes de Illumina Nextera XT se someten a la preparación de bibliotecas siguiendo el protocolo de bibliotecas de secuenciación metagenómica de Illumina de 16S. La secuenciación profunda se realiza en una plataforma Illumina Mi-Seq. Típicamente, se generan 200.000 lecturas de extremos emparejados de 250 pb (2 x 100.000 lecturas) por amplicón. Las lecturas se analizan utilizando CRISPResso (Pinello *et al.*, 2016) para calcular las tasas de edición. Los alineamientos de salida se seleccionan manualmente para confirmar los sitios de inserción y delección, así como para identificar sitios de microhomología en los sitios de recombinación.

Tabla 15: Sitios de unión de los cebadores de NGS

Dirección	Secuencia	SEQ ID NO.
Directo	5'-ACTTCCTCCAGCC-3'	223
Inversa	5'-GAACCCCGGCTTA-3'	224

Se realiza una transferencia de Western del lisado celular de células transfectadas y de células de control usando un anticuerpo anti-IDUA para verificar la expresión de la proteína de longitud completa y un ensayo de actividad enzimática en el lisado celular usando el sustrato 4-metilumbeliferil α -L-idurónido verifica que la enzima sea catalíticamente activa (Hopwood *et al.*, Clin.Chim. Acta (1979), 92(2): 257-265, que se incorporan como referencia en el presente documento). Estos experimentos se realizan en comparación con la línea celular con Idua^{W402X/W402X} original (sin transfección), la línea celular con Idua^{W402X/W402X} transfectada con la construcción de edición de bases y una secuencia guía aleatoria, y una línea celular que expresa IDUA de tipo silvestre.

Ejemplo 7. 4: Validación del tratamiento de enfermedades en un modelo murino

Para verificar la eficacia de este enfoque terapéutico, se utiliza un modelo de ratón con una mutación terminadora en el aminoácido análogo. La cepa de ratón porta una mutación W392X en su gen Idua (ID de Gene: 15932) que corresponde a la mutación homóloga en pacientes con síndrome de Hurler (Bunge *et al.*, (1994), Hum. Mol. Genet. 3(6): 861-866, que se incorporan como referencia en el presente documento). Este locus comprende una secuencia de nucleótidos distinta con respecto a la de seres humanos, que carece de la secuencia PAM necesaria para la corrección con los editores de bases descritos en los ejemplos anteriores y, por tanto, necesita el diseño de una proteína de fusión distinta para realizar la corrección de nucleótidos. La mejora de la enfermedad en este animal puede validar el enfoque terapéutico de corregir la mutación en tejidos accesibles mediante un vector de administración de genes.

Los ratones homocigóticos para esta mutación presentan una serie de características fenotípicas similares a las de los pacientes con síndrome de Hurler. Una proteína de fusión de edición de bases-NGA como se describe anteriormente (Tabla 12) junto con una secuencia guía de ARN se incorporan en un vector de expresión que permite la expresión de proteínas y la transcripción de ARN en ratones. A continuación en la Tabla 16 se muestra un diseño de estudio. El estudio incluye grupos que se tratan con una dosis alta del vector de expresión comprendiendo la proteína de fusión de edición de bases y la secuencia guía de ARN, una dosis baja del mismo vector de expresión, un control que es el ratón modelo tratado con un vector de expresión que no comprende la proteína de fusión de edición de bases ni el ARN de guía, y un segundo control que es un ratón de tipo silvestre tratado con el mismo vector vacío.

Tabla 16: Experimento de edición genómica en un modelo murino

Grupo	Cepa de ratón	N	Tratamiento
1	Idua-W392X ¹	≥ 5	Baja dosis de vector
2	Idua-W392X	≥ 5	Alta dosis de vector
3	Idua-W392X	≥ 5	Vehículo
4	129/Sv (TS)	5	Vehículo

Los criterios de valoración a evaluar incluyen el peso corporal, la excreción de GAG en orina, la actividad enzimática de IDUA sérica, la actividad de IDUA en tejidos de interés, la patología tisular, el genotipado de tejidos de interés para verificar la corrección del SNP y una evaluación del comportamiento y neurológica. Dado que algunos criterios de valoración son terminales, se pueden añadir grupos adicionales para la evaluación de, por ejemplo, la patología tisular y actividades tisulares de IDUA antes del final del estudio. Se pueden encontrar ejemplos adicionales de criterios de valoración en artículos publicados que establecen modelos animales para el síndrome de Hurler (Shull *et al.* (1994), Proc. Natl. Acad. Sci. U.S.A., 91(26): 12937-12941; Wang *et al.* (2010), Mol. Genet. Metab., 99(1): 62-71; Hartung *et al.* (2004), Mol. Ther., 9(6): 866-875; Liu *et al.* (2005), Mol. Ther., 11(1): 35-47; Clarke *et al.* (1997), Hum. Mol. Genet. 6(4): 503-511; todos ellos incorporados por referencia en el presente documento).

Un posible vector de administración utiliza el virus adenoasociado (AAV). Se produce un vector que incluya una secuencia codificante de proteína de fusión editor de bases-dNGA (por ejemplo, Nuc-ADAT-Enlazador-dAPG19748-Enlazador-SV40, como se describe anteriormente) precedida por un potenciador de CMV (SEQ ID NO: 259) y un promotor (SEQ ID NO: 258), u otra combinación adecuada de potenciador y promotor), opcionalmente una secuencia de Kozak, y fusionada operativamente en el extremo 3' a una secuencia terminadora y una secuencia de poliadenilación tal como la secuencia mínima descrita en Levitt, N.; Briggs, D.; Gil, A.; Proudfoot, N. J. Definition of an Efficient Synthetic Poly(A) Site. Genes Dev. 1989, 3 (7), 1019-1025. El vector puede comprender además un casete de expresión que codifica un ARN de guía único unido operativamente en su extremo 5' a un promotor de U6 humano (SEQ ID NO: 260) u otro promotor adecuado para la producción de pequeños ARN no codificantes, y que además comprenda secuencias de repetición terminal (ITR) necesarias y bien conocidas en la técnica para el empaquetamiento en la cápside del AAV. La producción y el empaquetamiento vírico se realizan mediante métodos convencionales, tales como los descritos en la patente de EE. UU. n.º 9.587.250, incorporados en el presente documento por referencia.

Otros posibles vectores víricos incluyen vectores de adenovirus y lentivirus, que se utilizan comúnmente y que contendrían elementos similares, con diferentes capacidades y requisitos de empaquetamiento. También se pueden utilizar métodos de administración no víricos, tales como ARNm y ARNgu encapsulados por nanopartículas lipídicas (Cullis, P. R. y Allen, T. M. (2013), Adv. Drug Deliv. Rev. 65(1): 36-48; Finn *et al.* (2018), Cell Rep. 22(9): 2227-2235, ambos incorporados por referencia) inyección hidrodinámica de ADN plasmídico (Suda T y Liu D,) 2007) Mol. Ther. 15(12): 2063-2069, incorporados en el presente documento por referencia) o complejos ribonucleoproteicos de ARNgu y asociados con nanopartículas de oro (Lee, K.; Conboy, M.; Park, H. M.; Jiang, F.; Kim, H. J.; Dewitt, M. A.; Mackley, V. A.; Chang, K.; Rao, A.; Skinner, C.; *et al.* Nanoparticle Delivery of Cas9 Ribonucleoprotein and Donor DNA *in Vivo* Induces Homology-Directed DNA Repair. Nat. Biomed. Eng. 2017, 1 (11), 889-90).

Ejemplo 7.5: Corrección de enfermedades en un modelo murino con un locus humanizado

Para evaluar la eficacia de una construcción de editor base idéntica a la que se usaría para la terapia humana, se necesita un modelo de ratón en el que los nucleótidos cercanos a W392 se alteren para que coincidan con la secuencia en seres humanos alrededor de W402. Esto se puede lograr mediante una diversidad de técnicas, incluido el uso de una NGA y un molde de RDH para cortar y reemplazar el locus en embriones de ratón.

Debido al alto grado de conservación de los aminoácidos, la mayoría de los nucleótidos en el locus de ratón pueden alterarse a los de la secuencia humana con mutaciones silenciosas, como se muestra en la Tabla 17. Los únicos cambios de base que dan como resultado una secuencia codificante alterada en el genoma de ratón técnicamente diseñado resultante se producen después del codón de parada introducido.

Tabla 17: Mutaciones de nucleótidos para generar un locus de ratón humanizado

Característica	Ser humano (W402X)		Ratón (W392X)		Ratón humanizado	
	Nucleótido (SEQ ID NO: 225)	AA codificado	Nucleótido (SEQ ID NO: 226)	AA codificado	Nucleótido (SEQ ID NO: 227)	AA codificado
	Ser humano (W402X)		Ratón (W392X)		Ratón humanizado	
Protoespaciador	G	E	A	G	G	G
	G	E	G	E	G	E
	A		A		A	
	G		A		G	
	C	Q	C	Q	C	Q
	A		A		A	

(continuación)

	Ser humano (W402X)		Ratón (W392X)		Ratón humanizado	
Característica	Nucleótido (SEQ ID NO: 225)	AA codificado	Nucleótido (SEQ ID NO: 226)	AA codificado	Nucleótido (SEQ ID NO: 227)	AA codificado
	Ser humano (W402X)		Ratón (W392X)		Ratón humanizado	
Protoespaciador	G		A		G	
	C	L	C	L	C	L
	T		T		T	
	C		C		C	
	T	PARADA	T	PARADA	T	PARADA
	A		A		A	
	G		G		G	
	G	A	G	A	G	A
	C		C		C	
	C		A		C	
	G	E	G	E	G	E
	A		A		A	
	A		G		A	
	G	V	G	V	G	V
	T		T		T	
	G		C		G	
	T	S	T	S	T	S
	C		C		C	
	G		A		G	
PAM, no crítico	C	Q	A	K	C	Q
	A		A		A	
	G		G		G	
	G	A	G	A	G	A
PAM, crítico	C		C		C	
	C		T		C	

Tras el diseño técnico de esta cepa de ratón, se realizarán experimentos similares a los que se describes en el Ejemplo 7.4.

5

Ejemplo 8: Direccionamiento a mutaciones responsables de la ataxia de Friedreich

La expansión de la secuencia de repetición trinucleotídica que provoca la ataxia de Friedreich (FRDA) de produce en un locus genético definido dentro del gen FXN, conocida como la región de inestabilidad de FRDA. Se pueden usar nucleasas guiadas por ARN (las NGA) para escindir la región de inestabilidad en células de pacientes con FRDA. Este enfoque precisa 1) una NGA y una secuencia de ARN de guía que pueda programarse para dirigirse al alelo en el genoma humano; y 2) un enfoque de administración para la NGA y la secuencia de guía. Muchas nucleasas utilizadas para la edición genómica, tales como la comúnmente utilizada nucleasa Cas9 de *S. pyogenes* (SpCas9), son demasiado grandes para empaquetarse en vectores de virus adenoasociado (AAV), especialmente cuando se considera la longitud del gen de la SpCas9 y el ARN de guía además de otros elementos genéticos necesarios para casetes de expresión funcionales. Esto hace que sea poco probable un enfoque viable que utilice SpCas9.

10

15

Las nucleasas compactas guiadas por ARN de la invención, por ejemplo, APG03850, son excepcionalmente adecuadas para la escisión de la región de inestabilidad de FRDA. APG03850 tiene un requisito de PAM que se encuentra cerca de la región de inestabilidad de FRDA. Además, APG03850 puede empaquetarse en un vector AAV junto con un ARN guía. El empaquetamiento de dos ARN de guía posiblemente podría precisar un segundo vector, pero este enfoque aún se compara favorablemente con lo que se precisaría de una nucleasa más grande tal como SpCas9, que puede dividir la secuencia proteica entre dos vectores.

La Tabla 18 muestra la ubicación de secuencias genómicas diana adecuadas para dirigir a APG03850 a los flancos 5' y 3' de la región de inestabilidad de FRDA, así como la secuencia de los ARN_{gu} para las dianas genómicas. Una vez en el locus, la NGA eliminaría la región de inestabilidad de AF. La eliminación de la región se puede verificar con la secuenciación del locus con Illumina.

Tabla 18: Secuencias genómicas diana para sistemas de NGA

Guía N.º	Ubicación relativa con respecto a la región de inestabilidad de FRDA	Secuencia diana genómica (SEQ ID NO:)	ARN _{gu} (SEQ ID NO:)
1	5'	228	229
2	5'	230	231
3	3'	232	233
4	3'	234	235

Ejemplo 9: Direccionamiento a mutaciones responsables de anemias drepanocíticas

Las secuencias dirigidas dentro de la región potenciadora de BCL11A (SEQ ID NO: 236) pueden proporcionar un mecanismo para aumentar la hemoglobina fetal (HbF) para curar o aliviar los síntomas de las enfermedades drepanocíticas. Por ejemplo, los estudios hologenómicos han identificado un conjunto de variaciones genéticas en BCL11A que están asociadas con niveles aumentados de HbF. Estas variaciones son una colección de SNP que se encuentran en regiones no codificantes de BCL11A que funcionan como una región potenciadora específica de fase y restringida al linaje. Una investigación adicional reveló que este potenciador de BCL11A es necesario en las células eritroides para la expresión de BCL11A (Bauer *et al.*, (2013) Science 343:253-257, que se incorporan como referencia en el presente documento). La región potenciadora se encontró dentro del intrón 2 del gen BCL11A y se identificaron tres áreas de hipersensibilidad a la ADNasa (a menudo indicativo de un estado de la cromatina asociado con el potencial regulador) en el intrón 2. Estas tres áreas se identificaron como "+62", "+58" y "+55" de conformidad con la distancia en kilobases desde el sitio de iniciación de la transcripción de BCL11A. Estas regiones potenciadoras son aproximadamente 350 (+55); 550 (+58); y 350 (+62) nucleótidos de longitud (Bauer *et al.*, 2013).

Ejemplo 9.1: Identificación de sistemas de NGA preferidos

En este caso se describe un posible tratamiento para las beta-hemoglobinopatías utilizando un sistema de NGA que altera la unión de BCL11A a su sitio de unión dentro del locus de HBB, que es el gen responsable de la producción de beta-globina en hemoglobina adulta. Este enfoque utiliza NHEJ, que es más eficaz en células de mamífero. Además, este enfoque utiliza una nucleasa de tamaño suficientemente pequeño que puede empaquetarse en un único vector de AAV el administración *in vivo*.

El motivo potenciador GATA1 en la región potenciadora de BCL11A humana (SEQ ID NO: 236) es una diana ideal para la alteración usando nucleasas guiadas por ARN (las NGA) para reducir la expresión de BCL11A con reexpresión simultánea de HbF en eritrocitos humanos de adulto (Wu *et al.* (2019) Nat Med 387:2554). Varias secuencias PAM compatibles con APG03850 o APG09208 son fácilmente evidentes en el locus genético que rodea este sitio GATA1. Estas nucleasas tienen una secuencia PAM de 5'-nnnnG-3' (SEQ ID NO: 42) y 5'-nnnnC-3' (SEQ ID NO: 62), respectivamente, y son de tamaño compacto, permitiendo posiblemente su administración junto con un ARN de guía apropiado en un único vector de AAV o adenovírico. Este enfoque de administración otorga múltiples ventajas en relación con otros, tal como el acceso a células madre hematopoyéticas y un perfil de seguridad y técnicas de fabricación bien establecidos.

La nucleasa Cas9 de *S. pyogenes* (SpyCas9) usada habitualmente precisa una secuencia PAM de 5'-NGG-3', (SEQ ID NO: 256) varias de las cuales están presentes cerca del motivo GATA1. Sin embargo, el tamaño de SpyCas9 impide el empaquetamiento en un único vector de AAV o adenovíricos y, por tanto, se pierden las ventajas antes mencionadas de este enfoque. Si bien se puede emplear una estrategia de administración doble, se añadiría una complejidad y un costo de fabricación significativos. Además, la administración de un vector vírico doble disminuye significativamente la eficacia de la corrección génica, ya que una edición satisfactoria en una célula determinada precisa la infección con ambos vectores.

Se produce un casete de expresión que codifica un APG03850 o APG09208 humano optimizado para el codón, similar a los descritos en el Ejemplo 4. También se producen casetes de expresión que expresan ARN de guía para las NGA

APG03850 o APG09208. Estos ARN de guía comprenden: 1) una secuencia protoespaciadora que es complementaria a la cadena de ADN codificante o no codificante dentro del locus del potenciador de BCL11A (la secuencia diana) y 2) una secuencia de ARN necesaria para la asociación del ARN de guía con la NGA. Debido a que varias posibles secuencias PAM para APG03850 o APG09208 rodean el motivo potenciador de BCL11A GATA1, se producen varias posibles construcciones de ARN de guía para determinar la mejor secuencia protoespaciadora que produzca una escisión robusta y una alteración mediada por NHEJ de la secuencia potenciadora de BCL11A GATA1. Las secuencias genómicas diana de la Tabla 19 se evalúan en cuanto a la dirección de la NGA a este locus usando el ARN_g proporcionado en la Tabla 19.

Tabla 19: Secuencias diana para el locus del potenciador GATA1 de BCL11A utilizando APG03850 o APG09208

Guía	NGA	Secuencia genómica diana (SEQ ID NO:)	ARN _g (SEQ ID NO:)
1	APG03850	237	238
2	APG03850	239	240
3	APG03850	241	242
4	APG09208	243	244
5	APG09208	245	246
6	APG09208	247	248

Para evaluar la eficacia con la que APG03850 o APG09208 generan inserciones o deleciones que alteran la región potenciadora de BCL11A, se utilizan líneas celulares humanas tales como células de riñón embrionario humano (células HEK). Se produce un vector de ADN que comprende un casete de expresión de NGA (por ejemplo, como se describe en el Ejemplo 4). También se produce un vector separado que comprende un casete de expresión que comprende una secuencia codificante para una secuencia de ARN de guía de la Tabla 16. Dicho casete de expresión puede comprender además un promotor de U6 de la ARN polimerasa III humana (SEQ ID NO: 260), como se describe en el Ejemplo 4. Como alternativa, se puede utilizar un único vector que comprende casetes de expresión tanto de la NGA como del ARN de guía. El vector se introduce en células HEK usando técnicas convencionales tales como las descritas en el Ejemplo 4, y las células se cultivan durante 1-3 días. Después de este período de cultivo, se aísla el ADN genómico y se determina la frecuencia de las inserciones o deleciones mediante digestión con endonucleasa I de T7 y/o secuenciación directa del ADN, como se describe en el Ejemplo 4.

Una región de ADN que abarca la región de BCL11A diana se amplifica mediante PCR con cebadores que contienen secuencias protuberantes de Illumina Nextera XT. Estos amplicones de PCR se examinan para determinar la formación de NHEJ mediante digestión con endonucleasa I de T7 o se someten a una preparación de bibliotecas siguiendo el protocolo de bibliotecas de secuenciación metagenómica de Illumina de 16S o una preparación de bibliotecas de secuenciación de próxima generación (NGS, del inglés *Next Generation Sequencing*) similar. Después de una secuenciación profunda, las lecturas generadas se analizan mediante CRISPResso para calcular las tasas de edición. Los alineamientos de salida se seleccionan manualmente para confirmar los sitios de inserción y deleción. Este análisis identifica la NGA preferida y el correspondiente ARN de guía (ARN_g) preferido. El análisis puede dar como resultado que tanto APG03850 como APG09208 sean igualmente preferidas. Además, el análisis puede determinar que hay más de un ARN de guía preferido, o que todas las secuencias genómicas diana en la Tabla 16 son igualmente preferidas.

Ejemplo 9.2: Ensayo para la expresión de hemoglobina fetal

En este ejemplo, las inserciones o deleciones generadas por APG03850 o APG09208 que alteran la región potenciadora de BCL11A se someten a ensayo en cuanto a la expresión de hemoglobina fetal. Se utilizan células madre hematopoyéticas (CMH) CD34⁺ de donante humano sano. Estas CMH se cultivan y se introduce un vector o vectores que comprenden casetes de expresión que comprenden las regiones codificantes de la NGA preferida y el ARN_g preferido usando métodos similares a los descritos en el Ejemplo 8.1. Después de la electroporación, estas células se diferencian *in vitro* en eritrocitos utilizando protocolos establecidos (por ejemplo, Giarratana *et al.* (2004) Nat Biotechnology 23: 69-74, incorporada en el presente documento por referencia). A continuación, la expresión de HbF se mide mediante transferencia de western con un anticuerpo anti-HbF humana o se cuantifica mediante cromatografía líquida de alta resolución (HPLC). Se espera que la alteración satisfactoria del locus potenciador de BCL11A conduzca a un aumento en la producción de HbF en comparación con las HSC electroporadas solo con la NGA pero sin guía.

Ejemplo 9.3: Ensayo para disminuir la formación de células falciformes

En este ejemplo, las inserciones o deleciones generadas por APG03850 o APG09208 que alteran la región potenciadora de BCL11A se someten a ensayo en cuanto a la formación de células falciformes. Se utilizan células

madre hematopoyéticas (CMH) CD34⁺ donantes de pacientes afectados por anemia drepanocítica. Estas CMH se cultivan y se introduce un vector o vectores que comprenden casetes de expresión que comprenden las regiones codificantes de la NGA preferida, y el ARN_g preferido usando métodos similares a los descritos en el Ejemplo 8.1. Después de la electroporación, estas células se diferencian *in vitro* en eritrocitos utilizando protocolos establecidos (Giarratana *et al.* (2004) Nat Biotechnology 23:69-74). A continuación, la expresión de HbF se mide mediante transferencia de western con un anticuerpo anti-HbF humana o se cuantifica mediante cromatografía líquida de alta resolución (HPLC). Se espera que la alteración satisfactoria del locus potenciador de BCL11A conduzca a un aumento en la producción de HbF en comparación con las HSC electroporadas solo con la NGA pero sin guía.

La formación de células falciformes se induce en estos eritrocitos diferenciados mediante la adición de metabisulfito. El número de eritrocitos falciformes frente a los normales se cuenta utilizando un microscopio. Se espera que el número de células falciformes sea menor en las células tratadas con APG03850 o APG09208 más los ARN_g que en las células no tratadas o las tratadas con NGA solo.

Ejemplo 9.4: Validación del tratamiento de enfermedades en un modelo murino

Para evaluar la eficacia del uso de la alteración por APG03850 o APG09208 del locus de BCL11A, se utilizan modelos de ratón humanizados adecuados de anemia falciforme. Los casetes de expresión que codifican la NGA preferida y el ARN_g preferido se empaquetan en vectores de AAV o vectores de adenovirus. En particular, el adenovirus tipo Ad5/35 es eficaz para dirigirse a las CMH. Se elige un modelo de ratón adecuado que contenga un locus HBB humanizado con alelos falciformes, tales como B6;FVB-Tg(LCR-HBA2,LCR-HBB*E26K)53Hhb/J o B6.Cg-Hb^{bm1Paz}Hbb^{tm1Tow} Tg(HBA-HBBs)41Paz/HhbJ. Estos ratones se tratan con factor estimulante de colonias de granulocitos solo o en combinación con plerixafor para movilizar las CMH a la circulación. Luego se inyectan por vía intravenosa los AAV o los adenovirus que portan la NGA y el plásmido del guía y se deja que los ratones se recuperen durante una semana. La sangre obtenida de estos ratones se analiza en un ensayo *in vitro* de falciformación utilizando metabisulfito, y los ratones se siguen longitudinalmente para controlar las tasas de mortalidad y la función hematopoyética. Se espera que el tratamiento con AAV o adenovirus que portan una NGA y un ARN de guía reduzca la falciformación, la mortalidad y mejore la función hematopoyética en comparación con los ratones tratados con virus que carecen de ambos casetes de expresión, o con virus que portan solo el casete de expresión de NGA.

Ejemplo 10: Ensayo de distintos formatos de administración

Para determinar si las NGA pueden administrarse en distintos formatos, se probó la nucleofección de ARNm y RNP con linfocitos T primarios. Se descongelaron linfocitos T CD3⁺ purificados o células mononucleares de sangre periférica (PBMC), activados utilizando microesferas CD3/CD28 (ThermoFisher) durante 3 días, a continuación se nucleofectaron utilizando la unidad Lonza 4D-Nucleofector X y las tiras Nucleocuvette. El kit P3 Primary Cell se utilizó tanto para la administración de ARNm como de RNP. Las células se transfectaron utilizando los programas EO-115 y EH-115 para la administración de ARNm y RNP respectivamente. Las células se cultivaron en medio de expansión de linfocitos T CTS OpTimizer (ThermoFisher) que contenía IL-2, IL-7 e IL-15 (Miltenyi Biotec) durante 4 días después de la nucleofección antes de ser cosechadas utilizando un kit de aislamiento de ADN genómico de tejido Nucleospin (Machery Nagel).

Los amplicones que rodean los sitios de edición se generaron mediante PCR utilizando los cebadores identificados en la Tabla 4 y se sometieron a secuenciación NGS mediante la plataforma Illumina Nextera utilizando secuenciación de extremos emparejados de 2x250 pb siguiendo el método del Ejemplo 4.

Tabla 20: Administración de ARNm y RNP de NGA en linfocitos T primarios

NGA	Método de liberación	SGN	% Edición
APG01604	ARNm, Nucleofección	SGN001671	1,18
APG01604	ARNm, Nucleofección	SGN001673	22,7
APG01604	RNP, Nucleofección	SGN001673	0,56
APG01604	ARNm, Nucleofección	SGN001674	0,63
APG01604	RNP, Nucleofección	SGN001674	0,77
APG01868	ARNm, Nucleofección	SGN001684	77,85
APG01868	RNP, Nucleofección	SGN001684	83,73
APG01868	ARNm, Nucleofección	SGN001691	82,53
APG01868	RNP, Nucleofección	SGN001691	80,7

(continuación)

NGA	Método de liberación	SGN	% Edición
APG01868	ARNm, Nucleofección	SGN001692	76,84
APG01868	RNP, Nucleofección	SGN001692	86,06
APG01868	ARNm, Nucleofección	SGN001785	0,22
APG01868	RNP, Nucleofección	SGN001785	0,43

Tanto la administración de ARNm como de RNP mostraron una edición con éxito con APG01868. APG01868 mostró una edición con éxito con RNP en varias dianas genómicas. La administración de RNP se limitó con APG01604, pero mostraron tasas de edición iguales con la administración de ARNm.

Ejemplo 11: Pruebas del Editor Base

Para determinar si APG09298 y APG01604 podían realizar la edición de bases de citosina en células de mamífero, se fusionó operativamente una citosina desaminasa con la versión nickasa de cada NGA para producir una proteína de fusión. Se identificaron los restos predichos para desactivar el dominio RuvC de las NGA APG09298 y APG01604 y se modificaron las NGA a variantes de nickasa. Una variante nickasa de una NGA se denomina en el presente documento "NGAn". Debe entenderse que puede utilizarse cualquier variante de nickasa de una NGA para producir una proteína de fusión de la invención.

Las secuencias de nucleótidos de la desaminasa y la NGAn optimizadas en codones para la expresión en mamíferos se sintetizaron como proteínas de fusión con un marcador en N-terminal de localización nuclear y se clonaron en el plásmido de expresión pTwist CMV (Twist Biosciences). Cada proteína de fusión comprende, a partir del extremo amino, la NLS de SV40 (SEQ ID NO: 251) unida operativamente en el extremo C-terminal al marcador 3X FLAG (SEQ ID NO: 252), unido operativamente en el extremo C-terminal a una desaminasa (APG05840), unida operativamente en el extremo C-terminal a un enlazador peptídico, unido operativamente en el extremo C-terminal a la NGAn (nAPG09298 o nAPG01604), unida operativamente en el extremo C-terminal a un enlazador peptídico, unido operativamente en el extremo C-terminal a una proteína estabilizadora del uracilo (USP2 establecida como SEQ ID NO: 1089), finalmente unida operativamente en el extremo C-terminal a la NLS de la nucleoplasmina (SEQ ID NO: 253). La secuencia de aminoácidos de las proteínas de fusión APG05840-nAPG09298-USP2 y APG05840-nAPG01604-USP2 se establece como las SEQ ID NO: 1090 y 1091, respectivamente.

También se produjeron plásmidos de expresión que incluían un casete de expresión que codificaba un ARN_g. Las secuencias genómicas humanas diana y las secuencias de ARN_g para guiar las proteínas de fusión a las dianas genómicas se indican en la Tabla 3 y los cebadores para la amplificación de la región genómica se enumeran en la Tabla 4. Los mismos métodos del Ejemplo 4 para la administración de plásmidos a células de mamíferos y la secuenciación de amplicones se utilizaron para probar las capacidades de edición de bases de estas NGA cuando se unen a una citosina desaminasa.

Tabla 21: Estimación de las tasas de edición de bases para cada NGA examinada

Construcción	Diana	% Lecturas mutadas
APG05840-nAPG09298-USP2	SGN001159	6,96
APG05840-nAPG09298-USP2	SGN001159	9,91
APG05840-nAPG09298-USP2	SGN001162	21,59
APG05840-nAPG09298-USP2	SGN001162	27,05
APG05840-nAPG09298-USP2	SGN001163	17,85
APG05840-nAPG09298-USP2	SGN001163	20,74
APG05840-nAPG09298-USP2	SGN001164	4,74
APG05840-nAPG09298-USP2	SGN001164	7,89
APG05840-nAPG09298-USP2	SGN001165	31,99
APG05840-nAPG09298-USP2	SGN001165	62,47
APG05840-nAPG09298-USP2	SGN001166	23,49

(continuación)

Construcción	Diana	% Lecturas mutadas
APG05840-nAPG09298-USP2	SGN001166	29,56
APG05840-nAPG01604-USP2	SGN001245	9,67
APG05840-nAPG01604-USP2	SGN001246	1,4
APG05840-nAPG01604-USP2	SGN001249	1,72
APG05840-nAPG01604-USP2	SGN001250	13,41
APG05840-nAPG01604-USP2	SGN001251	0,87
APG05840-nAPG01604-USP2	SGN001252	4,28

Ejemplo 12: Análisis de inespecificidad

- 5 Para evaluar la especificidad de las nucleasas, la edición no específica se determinó en sitios potenciales identificados mediante bioinformática. Los posibles sitios no específicos para APG01604 se identificaron por dianas con menos de cinco desajustes en la secuencia diana y al menos un resto coincidente en la secuencia PAM.

- 10 Se utilizaron los mismos métodos que los descritos en el Ejemplo 4 para la administración de plásmidos a células de mamífero y la secuenciación de amplicones para probar la especificidad y la edición fuera de diana de APG01604. Del mismo experimento en el que se probaron los SGN del cuadro 22 para la edición en el objetivo, las posibles ubicaciones fuera del objetivo de la Tabla 23 se ensayaron para su posible edición. Los cebadores de la Tabla 24 se utilizaron para amplificar posibles sitios fuera de diana con similitud de secuencia con el sitio en diana para buscar la edición fuera de diana.

15

Tabla 22: SGN usados para buscar ediciones no específicas

SGN	Gen	Cebador directo en la diana (SEQ ID NO)	Cebador Inverso en la diana (SEQ ID NO)
SGN001675	TRA	CCCTTGTCATCACTGGCAT (1096)	ACCAAAGCTGCCCTTACCTG (1097)
SGN001594	B2M	CCTTAATGTGCCTCCAGCCT (1092)	AGGAGAGACTCACGCTGGAT (1093)
SGN001674	TRA	CCCTTGTCATCACTGGCAT (1094)	ACCAAAGCTGCCCTTACCTG (1095)

Tabla 23: Secuencias no específicas analizadas

SGN	Número de no específicos	Secuencia de locus no específico	SEQ ID NO
SGN001594	1594-2	AGCACAGCTAAGGCCTAAAGTTGAA	1098
SGN001594	1594-3	AGCACAGCTAAGGCACCGATTGGA	1099
SGN001594	1594-4	AGCACAGCCAAGGCCAAGGGCTGAC	1100
SGN001594	1594-5	AGCAGAGCTAAGGCCAAGGCAGGTG	1101
SGN001594	1594-6	GGCACAGCTAAGGCCAGCAGTGGCC	1102
SGN001674	1674-2	GTCTCTGAGCTGGTACATGGCAGAG	1103
SGN001674	1674-3	GTCTTTTAGCTGGTACACGTGTGTC	1104
SGN001675	1675-2	AGAAGATTTGTCACTGGATTCTGAG	1105
SGN001675	1675-3	AGGACACTTGTCACTGGATTTAGGA	1106
SGN001675	1675-4	AGGACACTTGTCACTGGATTTAGGA	1107
SGN001675	1675-5	TGAGGACTTGTCACTGGATTCAGGG	1108
SGN001675	1675-6	AGGAGACTTTTCACTGGATTTAGGG	1109
SGN001675	1675-7	TGGACACTTGTCACTGGATTTAGGG	1110
SGN001675	1675-8	CACAGACTTGTCACTGGATGTGGGG	1111

(continuación)

SGN	Número de no específicos	Secuencia de locus no específico	SEQ ID NO
SGN001675	1675-9	TACAGACATGTCCTGGATCTGGAA	1112
SGN001675	1675-10	AACACACTTGTCTTGGATTAGGG	1113
SGN001675	1675-11	TACAGACTGGTCACTGGATGCTGGT	1114
SGN001675	1675-12	AGGACACTTGTCACTGGATTAGGA	1115
SGN001675	1675-13	CTAAGACTTGTTACTGGATTGTGTG	1116
SGN001675	1675-14	CTCAGACTGGTCACTGGATAATGTA	1117

Tabla 24: Cebadores usados para amplificar regiones no específicas

Descripción	Secuencia de cebador	SEQ ID NO
1594-2 DIR	CCAGAAGCCAGCAGATGACA	1118
1594-2 INV	GAGTGGTGGGCTCTCAATCC	1119
1594-3 DIR	AGCAACACCATCCAAAGGTT	1120
1594-3 INV	GGGAGTGATGATAATGCGGG	1121
1594-4 DIR	TGGACTAGAGAGGGTTGGGG	1122
1594-4 INV	TCTCTTTCCACGAGCAGCAG	1123
1594-5 DIR	GCCTTTGACCTTCCCAGATT	1124
1594-5 INV	ACCATTGGAAAGGTGGATGC	1125
1594-6 DIR	GGCTTCAGGCTTTCCTCTGT	1126
1594-6 INV	AGCATGCTGGCCTAAAGTGA	1127
1674-2 DIR	ACCATTGGTCTGCTCAGGTG	1128
1674-2 INV	CCAAAACCTGCAGTGGCTTC	1129
1674-3 DIR	GGAGAGGAACTGGGCATGAG	1130
1674-3 INV	TCCGTCTCTCCTAGGTCTGC	1131
1675-2 DIR	CCATGACTGGCCCTTCTGTT	1132
1675-2 INV	GGGTAGAGTACATGGCGACG	1133
1675-3 DIR	CCCTCCCACCAGAAGCTCTA	1134
1675-3 INV	GATAAGAGGCCCAAGGACCG	1135
1675-4 DIR	CACTTGACACGTGAGCCTCT	1136
1675-4 INV	CCTCTTTCAGCCTCTGGTGG	1137
1675-5 DIR	GACAGACTTGGTTCTGCCCT	1138
1675-5 INV	CCTCCCCTCCTCCTAGCTT	1139
1675-6 DIR	TGCTGTAGTGGGTGGAACAG	1140
1675-6 INV	TTTCACCATGCTGGCTAGGC	1141
1675-7 DIR	CCCAGCCACATGGAACCTAAG	1142
1675-7 INV	GGACCCTAGGAGTTCCTTGT	1143
1675-8 DIR	GCAGGTTGATAGGGAAGAGC	1144
1675-8 INV	TCATCTCCCAGCTGATGACA	1145

(continuación)

Descripción	Secuencia de cebador	SEQ ID NO
1675-9 DIR	CCACCTGTTGCACAAATCCG	1146
1675-9 INV	TCCTCCCCTGGGAACATGAT	1147
1675-10 DIR	GAGGTACTGGGAGTGGGGAT	1148
1675-10 INV	CTTCCTGCCTCTTCCAGCTC	1149
1675-11 DIR	GCACAGCTTTTGTTCATGGGG	1150
1675-11 INV	GGGGATGAGAAAACAGAGCCA	1151
1675-12 DIR	CCATGCCCCATTCTGAAGGT	1152
1675-12 INV	TTCCCCTCCTCTAGACTGCC	1153
1675-13 DIR	CTTGAGCCCAGGAGTTTGAG	1154
1675-13 INV	TGCATTCTTGGGATGACCTC	1155
1675-14 DIR	ACTGCTTTTCCCTGGACACA	1156
1675-14 INV	GGTCACAAGTCCCACTGGTC	1157

Utilizando la administración de plásmidos, no se detectó edición no específica en dos de las tres guías probadas. Un lugar no específico para SGN001675, 1675-8, mostró un 2 % de edición en el locus no específico. Hubo una duplicación de 132 pb de la región genómica en el sitio de corte en las lecturas de NGS. Dos sitios no específicos, 1675-7 y 1675-13, no pudieron secuenciarse debido a una amplificación deficiente de los cebadores.

Tabla 25: Edición no específica para SGN001594 y APG08167

Número de no específicos	Desajustes totales en la diana	Desajustes totales en la siembra	Coincidencias totales en la APM	% Edición
SGN001594	0	0	3	42,79
1594-2	2	2	3	0
1594-3	2	2	3	0,01
1594-4	2	1	2	0
1594-5	3	2	2	0
1594-6	2	1	1	0

Tabla 26: Edición no específica para SGN001675 y APG08167

Número de no específicos	Desajustes totales en la diana	Desajustes totales en la siembra	Coincidencias totales en la APM	% Edición
SGN001675	0	0	3	33,115
1675-8	2	1	3	2,02
1675-5	4	0	3	0,06
1675-10	3	1	3	0,03
1675-12	4	0	2	0,02
1675-6	4	0	3	0
1675-2	4	0	3	0
1675-3	4	0	3	0
1675-4	4	0	3	0
1675-9	3	1	3	0

(continuación)

Número de no específicos	Desajustes totales en la diana	Desajustes totales en la siembra	Coincidencias totales en la APM	% Edición
1675-11	2	0	2	0
1675-14	4	1	2	0
1675-13	4	0	2	Secuenciación problema
1675-7	4	0	3	Secuenciación problema

Tabla 27: Edición no específica para SGN001674 y APG08167

Número de no específicos	Desajustes totales en la diana	Desajustes totales en la siembra	Coincidencias totales en la APM	% Edición
SGN001674	0	0	3	52,17
1674-2	2	1	3	0
1674-3	2	0	1	0,03

5 **Ejemplo 13: Pruebas de variantes de la estructura principal del ARN guía**

APG08167 y APG01604 son aproximadamente un 62,72 % idénticos, pero reconocen la misma PAM, comparten el mismo ARNcr y tienen una secuencia de ARNtrat estrechamente relacionada. Debido a la similitud de las secuencias de ARN guía, todos los datos generados previamente para APG01604 utilizaron la estructura principal del ARNtrat APG08167. Estos estudios se realizaron para confirmar que la proteína APG01604 es más activa con un ARNtrat codificado en el genoma APG08167 (el ARNtrat APG08167 nativo). También se probaron diferentes longitudes de espaciador en el ARNcr para determinar si la proteína APG01604 prefiere una secuencia espaciadora de 20 o 25 pares de bases.

15 Para ello, se generaron ARNcr sintéticos con seis secuencias diana diferentes que contenían una diana de 20 o 25 pares de bases. Los diferentes ARNcr se combinaron con un ARNtrat sintético del genoma APG08167 o APG01604. Se formaron RNP y los complejos RNP se nucleofectaron en células HEK293T. Para determinar la tasa de edición en las células se utilizaron los métodos estándar del Ejemplo 4.

20

Tabla 28: Secuencias de ARNcr

Gen	Espaciador	Longitud del espaciador	SEQ ID NO
B2M	A	20	1158
	B	20	1159
	C	20	1160
	D	20	1161
	A	25	1162
	B	25	1163
	C	25	1164
	D	25	1165
TRA	E	20	1166
	F	20	1167
	E	25	1168
	F	25	1169

Tabla 29: Secuencias ARNtrat

ARNtrat	SEQ ID NO
ARNtrat APG08167	1170
ARNtrat APG01604	1171

Tabla 30: Secuencias de cebadores de secuenciación

Cebador	SEQ ID NO
Cebador izquierdo TRAC	1170
Cebador derecho TRAC	1171
Cebador izquierdo B2M	1172
Cebador derecho B2M	1173

Tabla 31: Resultados de la edición de la estructura principal de APG01604 y pruebas de longitud diana

Gen	Espaciador	Longitud del espaciador	Estructura principal	Eficiencia de la edición (%)
B2M	A	20	APG01604 tr	0,21
	B	20	APG01604 tr	0,08
	C	20	APG01604 tr	0,27
	D	20	APG01604 tr	0,17
	A	25	APG01604 tr	0,15
	B	25	APG01604 tr	0
	C	25	APG01604 tr	1,33
	D	25	APG01604 tr	0,31
	A	20	APG08167 tr	6,58
	B	20	APG08167 tr	1,75
	C	20	APG08167 tr	6,99
	D	20	APG08167 tr	64,6
	A	25	APG08167 tr	22,7
	B	25	APG08167 tr	2,36
	C	25	APG08167 tr	52,51
	D	25	APG08167 tr	82,58
TRA	E	20	APG01604 tr	0,05
	F	20	APG01604 tr	0,01
	E	25	APG01604 tr	0,02
	F	25	APG01604 tr	0,01
	E	20	APG08167 tr	0
	F	20	APG08167 tr	0,29
	E	25	APG08167 tr	2,64
	F	25	APG08167 tr	1,13

5

Las tasas de edición en células de mamíferos demostraron que no se producía una edición robusta cuando se utilizaba la secuencia nativa de ARNtrat APG01604. El RNP APG01604 mostró tasas de edición mucho más altas cuando se utilizó la secuencia de ARNtrat APG08167. Además, al comparar la misma diana, la longitud del espaciador de 25 pares de bases mostró una mayor edición en las dianas A, C, D, E y F que los ARNcr que tenían el espaciador de 20 pb de longitud. Estos resultados combinados demuestran que el ARNtrat APG08167 exhibe una mayor edición que el ARNtrat APG01604. Además, el RNP APG01604 funciona mejor con una secuencia espaciadora de 25 pb que con una de 20 pb.

10

REIVINDICACIONES

1. Una molécula de ácido nucleico que comprende un polinucleótido que codifica un polipéptido de nucleasa guiada por ARN (NGA), en donde dicho polinucleótido comprende una secuencia de nucleótidos que codifica un polipéptido de NGA que comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83;
5 en donde dicho polipéptido de NGA se puede unir a una secuencia de ADN diana de una manera específica de secuencia guiada por ARN cuando se une a un ARN de guía (ARNg) capaz de hibridar con dicha secuencia de ADN diana y
10 en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.
2. La molécula de ácido nucleico de la reivindicación 1, en donde la molécula de ácido nucleico tiene codones optimizados para su expresión en una célula eucariota.
3. La molécula de ácido nucleico de las reivindicaciones 1 o 2, en donde dicho polinucleótido que codifica un polipéptido NGA está operativamente unido a un promotor.
- 20 4. La molécula de ácido nucleico de la reivindicación 3, en donde dicho promotor es heterólogo para dicho polinucleótido.
5. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-3, en donde el polipéptido NGA está fusionado operativamente a un dominio efector.
- 25 6. La molécula de ácido nucleico de la reivindicación 5, en donde el dominio efector comprende un dominio de escisión, un dominio de desaminasa o un dominio modulador de expresión.
7. La molécula de ácido nucleico de la reivindicación 5 o la reivindicación 6, en donde el dominio efector está fusionado operativamente en el extremo N-terminal, el extremo C-terminal o en una localización interna de dicho polipéptido NGA.
- 30 8. Un vector que comprende la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-7.
- 35 9. El vector de la reivindicación 8, comprendiendo además al menos una secuencia de nucleótidos que codifica dicho ARNg capaz de hibridarse con dicha secuencia de ADN diana.
10. Una célula que comprende la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-7.
- 40 11. Un polipéptido aislado de nucleasa guiada por ARN (NGA), en donde dicho polipéptido NGA comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83; y
45 en donde dicho polipéptido NGA se puede unir a una secuencia de ADN diana de una molécula de ADN de una manera específica de secuencia guiada por ARN cuando se une a un ARN de guía (ARNg) capaz de hibridar con dicha secuencia de ADN diana; y
en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.
- 50 12. El polipéptido NGA de la reivindicación 11, en donde el polipéptido NGA está fusionado operativamente a un dominio efector.
13. El polipéptido NGA de la reivindicación 12, en donde el dominio efector comprende un dominio de escisión, un dominio de desaminasa o un dominio modulador de expresión.
- 55 14. El polipéptido NGA de la reivindicación 12 o la reivindicación 13, en donde el dominio efector está fusionado operativamente en el extremo N-terminal, el extremo C-terminal o en una localización interna de dicho polipéptido NGA.
- 60 15. Un sistema para unir una secuencia de ADN diana de una molécula de ADN, comprendiendo dicho sistema:
a) uno o más ARN de guía capaces de hibridar con dicha secuencia de ADN diana o uno o más polinucleótidos que comprenden una o más secuencias de nucleótidos que codifican el uno o más ARN de guía (ARNg); y
b) un polipéptido de nucleasa guiada por ARN (NGA) que comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83 o un polinucleótido que comprende una secuencia de nucleótidos que codifica el polipéptido de NGA;
65 en donde el uno o más ARN de guía pueden hibridar con la secuencia de ADN diana;

en donde el uno o más ARN guía son capaces de formar un complejo con el polipéptido NGA para dirigir dicho polipéptido NGA a unirse a dicha secuencia de ADN diana de la molécula de ADN; y

en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82.

16. El sistema de la reivindicación 15, en donde al menos una de dichas secuencias de nucleótidos que codifican el uno o más ARN guía y dicha secuencia de nucleótidos que codifica el polipéptido NGA está operativamente unida a un promotor heterólogo a dicha secuencia de nucleótidos.

17. El sistema de las reivindicaciones 15 o 16, en donde dicha secuencia de ADN diana es una secuencia de ADN diana eucariota.

18. El sistema de una cualquiera de las reivindicaciones 15-17, en donde las secuencias de nucleótidos que codifican el uno o más ARN guía y la secuencia de nucleótidos que codifica un polipéptido NGA se encuentran en un vector.

19. El sistema de una cualquiera de las reivindicaciones 15-18, en donde dicho ARNg se selecciona de:

a) un ARNg que comprenda una secuencia repetida CRISPR que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 84 y un ARNtrat que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 85; y

b) un ARNg que comprenda una secuencia repetida CRISPR que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 84 y un ARNtrat que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 78.

20. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-4, el polipéptido NGA aislado de la reivindicación 11 o el sistema de una cualquiera de las reivindicaciones 15-19, en donde dicho polipéptido NGA no tiene actividad de nucleasa o es una nickasa.

21. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-4 y 20, el polipéptido NGA aislado de las reivindicaciones 11 o 20 o el sistema de una cualquiera de las reivindicaciones 15-20, en donde el polipéptido de NGA está fusionado operativamente a un polipéptido de edición de bases.

22. La molécula de ácido nucleico, el polipéptido NGA aislado o el sistema de la reivindicación 21, en donde el polipéptido de edición de bases es una desaminasa.

23. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-4 y 20-22, el polipéptido NGA aislado de una cualquiera de las reivindicaciones 11 y 20-22 o el sistema de una cualquiera de las reivindicaciones 15-22, en donde el polipéptido NGA comprende una o más señales de localización nuclear.

24. Un método *in vitro* o *ex vivo* para unir una secuencia de ADN diana de una molécula de ADN que comprende administrar un sistema de acuerdo con una cualquiera de las reivindicaciones 15-19, a dicha secuencia de ADN diana o a una célula que comprende la secuencia de ADN diana, en donde el método no es un proceso para modificar la identidad genética de la línea germinal de un ser humano.

25. Un método *in vitro* o *ex vivo* para escindir y/o modificar una secuencia de ADN diana de una molécula de ADN, que comprende poner en contacto la molécula de ADN con:

a) un polipéptido de nucleasa guiada por ARN (NGA), en donde dicha NGA comprende una secuencia de aminoácidos que tiene al menos el 90 % de identidad de secuencia con la SEQ ID NO: 83; y

b) uno o más ARN de guía capaces de dirigir la NGA de (a) a la secuencia de ADN diana; en donde dicho polipéptido de NGA reconoce una secuencia de motivo adyacente al protoespaciador (PAM) que comprende la secuencia consenso de SEQ ID NO: 82;

en donde el uno o más ARN de guía hibridan con la secuencia de ADN diana, dirigiendo así dicho polipéptido de NGA para que se una a dicha secuencia de ADN diana y se produzca la escisión y/o modificación de dicha secuencia de ADN diana, y en donde el método no es un proceso para modificar la identidad genética de la línea germinal de un ser humano.

26. El método de la reivindicación 25, en donde la escisión por dicho polipéptido NGA genera una rotura bicatenaria o monocatenaria.

27. El método de las reivindicaciones 25 o 26, en donde dicha secuencia de ADN diana modificada comprende:

a) delección de al menos un nucleótido de la secuencia de ADN diana; o

b) mutación de al menos un nucleótido en la secuencia de ADN diana.

28. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-7, el vector de la reivindicación 8 o 9, la célula de la reivindicación 10, el polipéptido NGA aislado de una cualquiera de las reivindicaciones 11-14, el sistema de una cualquiera de las reivindicaciones 15-19 o el método de una cualquiera de las reivindicaciones 24-27, en donde dicho polipéptido NGA comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 83.