

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 699 848**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 15/113** (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **21.10.2016 PCT/US2016/058294**
- 87 Fecha y número de publicación internacional: **27.04.2017 WO17070598**
- 96 Fecha de presentación y número de la solicitud europea: **21.10.2016 E 16794469 (3)**
- 97 Fecha y número de publicación de la concesión europea: **26.09.2018 EP 3350327**

54 Título: **Ácido nucleico CRISPR clase 2 de tipo cruzado modificado que se dirige a ácidos nucleicos**

30 Prioridad:

**23.10.2015 US 201562245918 P**

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

**13.02.2019**

73 Titular/es:

**CARIBOU BIOSCIENCES, INC. (100.0%)  
2929 7th Street Suite 105  
Berkeley, CA 94710, US**

72 Inventor/es:

**DONOHOU, PAUL, DANIEL y  
MAY, ANDREW, PAUL**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

ES 2 699 848 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

Ácido nucleico CRISPR clase 2 de tipo cruzado modificado que se dirige a ácidos nucleicos

5 **Campo técnico**

La presente divulgación se refiere en general a ácidos nucleicos modificados que se dirigen a ácidos nucleicos y complejos nucleoproteicos que comprende una o más proteínas Cas y uno o más ácidos nucleicos modificados que se dirigen a ácidos nucleicos. La divulgación también se refiere a composiciones y métodos para fabricar y utilizar el ácido nucleico modificado que se dirige a ácidos nucleicos y complejos nucleoproteicos de la presente invención.

**Antecedentes**

Las repeticiones cortas palindrómicas interespaciadas agrupadas regularmente (CRISPR) y proteínas asociadas a CRISPR (Cas) constituyen el sistema CRISPR-Cas. El sistema CRISPR-Cas proporciona una inmunidad adaptativa contra el ADN ajeno en bacterias (véase, por ejemplo, Barrangou, R., et al, *Science* 315:1709-1712 (2007); Makarova, K. S., et al, *Nature Reviews Microbiology* 9:467-477 (2011); Garneau, J. E., et al, *Nature* 468:67-71 (2010); Sapranaukas, R., et al, *Nucleic Acids Research* 39:9275-9282 (2011)).

Los sistemas CRISPR-Cas se han reclasificado recientemente en dos clases, que comprenden cinco tipos y dieciséis subtipos (véase, Makarova, K., et al, *Nature Reviews Microbiology* 13:1-15 (2015)). Esta clasificación se basa en la identificación de todos los genes *cas* en un locus CRISPR-Cas y que determinan los genes de firma en cada locus CRISPR-Cas, determinando en último término que los sistemas CRISPR-Cas se pueden colocar en cualquiera de la Clase 1 o Clase 2 basándose en los genes que codifican el módulo efector, es decir, las proteínas implicadas en el estadio de interferencia. Recientemente se ha identificado un sexto sistema CRISPR-Cas (Tipo IV) (véase, Abudayyeh O., et al, *Science* 353(6299): aaf5573 (2016)). Ciertas bacterias poseen más de un tipo de sistema CRISPR-Cas.

Los sistemas de Clase 1 tienen un complejo multi-subunidad crARN-efectora, mientras que los sistemas Clase 2 tienen una única proteína, tal como Cas9, Cpf1, C2c1, C2c2, C2c3, o un complejo crARN-efector. Los sistemas Clase 1 comprenden los sistemas Tipo I, Tipo III, y Tipo IV. Los sistemas Clase 2 comprenden los sistemas Tipo II y Tipo V.

Los sistemas Tipo II tienen los genes *cas1*, *cas2* y *cas9*. El gen *cas9* codifica una proteína multi-dominio que combina las funciones del complejo crARN-efector con la escisión de la secuencia de ADN diana. Los sistemas Tipo II también codifican un tracrARN. Los sistemas Tipo II se dividen adicionalmente en tres subtipos, los subtipos II-A, II-B, y II-C. El subtipo II-A contiene un gen adicional *csn2*. Ejemplos de organismos con sistemas de subtipo II-A incluyen, pero no se limitan a, *Streptococcus pyogenes*, *Streptococcus thermophilus*, y *Staphylococcus aureus*. El subtipo II-B carece de *csn2*, por tiene el *cas4*. Un ejemplo de un organismo con un sistema de subtipo II-B es *Legionella pneumophila*. El subtipo II-C es el sistema de Tipo II más común que se encuentra en bacterias y tiene solo tres proteínas Cas1, Cas2 y Cas 9. Un ejemplo de un organismo con un sistema de subtipo II-C es *Neisseria lactamica*.

Los sistemas Tipo V tienen un gen *Cpf1* y genes *cas1* y *cas2* (véase, Zetsche, B., et al, *Cell* 163:1-13 (2015)). El gen *Cpf1* codifica una proteína Cpf1, que tiene un dominio nucleasa tipo-RuvC que es homólogo del dominio respectivo de Cas9, pero carece del dominio nucleasa HNH que está presente en la proteína Cas9. Los sistemas Tipo V se han identificado en varias bacterias, incluyendo, pero sin limitarse a, *Parcubacteria bacterium*, *Lachnospiraceae bacterium*, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium*, *Acidaminococcus spp.*, *Porphyromonas macacae*, *Porphyromonas crevioricanis*, *Prevotella disiens*, *Moraxella bovoculi*, *Smithella spp.*, *Leptospira inadai*, *Franciscella tularensis*, *Franciscella novicida*, *Candidatus methanoplasma termatum*, y *Eubacterium eligens*. Recientemente se ha demostrado que Cpf1 también tiene actividad RNasa, y es responsable del procesamiento de crARN (véase Fonfara, I., et al., *Nature* 532(7600):517-521 (2016)).

En los sistemas Clase 2, el crARN se asocia con una única proteína y consigue la interferencia combinando la actividad de nucleasa con dominios de unión al ARN y formación de pares de bases entre el crARN y una secuencia del ácido nucleico diana.

En los sistemas Tipo II, la secuencia de unión al ácido nucleico diana implica Cas9 y crARN, como secuencia de escisión del ácido nucleico diana. En los sistemas Tipo II, el dominio nucleasa tipo RuvC (RNasa con plegamiento en H) y el dominio nucleasa HNH (tipo McrA) de Cas9 escinde cada una, una de las cadenas de la secuencia del ácido nucleico diana de cadena doble. La actividad de escisión de la Cas9 de los sistemas Tipo II también necesitan la hibridación de crARN con tracrARN para formar un dúplex que facilita la unión del crARN y la secuencia del ácido nucleico diana por Cas9.

En los sistemas Tipo V, la secuencia de unión al ácido nucleico diana implica Cpf1 y el crARN, como secuencia de escisión del ácido nucleico diana. En los sistemas Tipo V, el dominio de nucleasa tipo RuvC de Cpf1 escinde una cadena de la secuencia del ácido nucleico diana de cadena doble, y un supuesto dominio de nucleasa escinde la otra cadena de la secuencia del ácido nucleico diana de cadena doble en una configuración escalonada, que produce protuberancias 5', que contrastan con los extremos romos generados por la escisión de Cas9. Estas protuberancias

5' pueden facilitar la inserción del ADN.

La actividad de escisión de CpfI de los sistemas Tipo V tampoco necesita la hibridación de crARN a tracrARN para formar un dúplex, más bien el crARN de los sistemas Tipo V utiliza un único crARN que tiene una estructura tallo-lazo que forma un dúplex interno. El CpfI se une al crARN de una manera específica de secuencia y estructura que reconoce el tallo-lazo y las secuencias adyacentes al tallo-lazo, sobre todo el nucleótido 5' de las secuencias espaciadoras que se hibridan a la secuencia del ácido nucleico diana. Esta estructura tallo-lazo normalmente está en el intervalo de 15 a 18 nucleótidos de longitud. Las sustituciones que alteran este dúplex tallo-lazo anulan la actividad de escisión, mientras que otras sustituciones que no alteran este dúplex tallo-lazo no anulan la actividad de escisión. En los sistemas Tipo V, el crARN forma una estructura de tallo-lazo en el extremo 5', y la secuencia del extremo 3' es complementaria a una secuencia de la secuencia del ácido nucleico diana.

Otras proteínas asociadas con el crARN del Tipo V y la secuencia de unión y escisión del ácido nucleico diana incluyen el candidato 1 de la Clase "(C2c1), y el candidato 3 de la Clase 2 (C2c3). Las proteínas C2c1 y C2c3 son similares en longitud a las proteínas Cas9 y CpfI, varían desde aproximadamente 1.100 aminoácidos a aproximadamente 1.500 aminoácidos. Las proteínas C2c1 y C2c3 también contienen dominios nucleasa tipo RuvC y tienen una arquitectura similar a CpfI. Las proteínas C2c1 son similares a las proteínas Cas9 porque necesitan un crARN y un tracrARN para la unión y escisión de la secuencia de ácido nucleico diana, pero tienen una temperatura óptima de escisión de 50 °C. Las proteínas C2c1 se dirigen a un motivo adyacente al protoespaciador rico en AT (PAM), similar al PAM de CpfI, que está 5' de la secuencia del ácido nucleico diana (véase, por ejemplo, Shmakov, S., et al, *Molecular Cell* 60(3):385-397 (2015)).

El candidato 2 de Clase 2 (C2c2) no comparte una similitud de secuencia con otras proteínas efectoras CRISPR y se identificó recientemente como un sistema Tipo VI (véase, Abudayyeh, O., et al, *Science* 353(6299): aaf5573 (2016)). Las proteínas C2c2 tienen dos dominios HEPN y se ha demostrado una actividad de escisión del ARN de cadena sencilla. Las proteínas C2c2 son similares a las proteínas CpfI porque necesitan un crARN para la unión y escisión de la secuencia del ácido nucleico diana, aunque no necesitan tracrARN. También al igual que CpfI, el crARN para las proteínas C2c2 forma una horquilla estable, o una estructura tallo-lazo, que ayuda a la asociación con la proteína C2c2. El Tipo VI es una ARN endonucleasa polipeptídica sencilla que utiliza un crARN sencillo para la escisión específica del sitio. Adicionalmente, después de la hibridación con el ARN diana complementario del espaciador, la C2c2 se convierte en una ARN endonucleasa promiscua que presenta una actividad endonucleasa no específica hacia cualquier ARN de cadena sencilla de una manera independiente de secuencia (véase, East-Seletsky, A., et al, *Nature* 538(7624):270-273 (2016)).

Con respecto a los sistemas CRISPR-Cas Clase 2 Tipo II, se conoce un gran número de ortólogos de Cas9 en la técnica así como sus componentes polinucleotídicos asociados (tracrARN y crARN) (véase, por ejemplo, Fonfara, I., et al, *Nucleic Acids Research* 42(4):2577-2590 (2014), incluyendo todos los Datos Suplementarios; Chylinski K., et al, *Nucleic Acids Research* 42(10):6091-6105 (2014), incluyendo todos los Datos Suplementarios). Además, se conocen proteínas sintéticas tipo Cas9 en la técnica (véase, la Solicitud de Patente Publicada de EE. UU. N.º 2014-0315985, publicada el 23 de octubre de 2014).

La Cas9 es una proteína CRISPR Cas Tipo II a modo de ejemplo. La Cas9 es una endonucleasa que se puede programar por el tracrARN/crARN para escindir, específicamente del sitio, una secuencia de ADN diana utilizando dos dominios de endonucleasa distintos (dominios HNH y tipo H RuvC/RNasa) (véase, la Solicitud de Patente Publicada de EE. UU. N.º 2014-0068797, publicada el 6 de marzo de 2014; véase también Jinek M., et al, *Science* 337:816-821 (2012)).

Normalmente, cada sistema CRISPR-Cas9 de tipo silvestre incluye un crARN y un tracrARN. El crARN tiene una región de complementariedad para una secuencia de ADN diana y una segunda región que forma enlaces hidrógeno entre pares de bases con el tracrARN para formar una estructura secundaria, normalmente para formar al menos una estructura entallo. La región de complementariedad de la secuencia del ADN diana es el espaciador. El tracrARN y crARN interactúan mediante varios enlaces de hidrógeno entre pares de bases para formar estructuras secundarias de ARN. La formación de complejos entre tracrARN/crARN y la proteína Cas9 da como resultado un cambio conformacional de la proteína Cas 9 que facilita la unión al ADN, las actividades de endonucleasa de la proteína Cas9, y la escisión de ADN específica del sitio guiada por crARN por la endonucleasa Cas9. Para que un complejo Cas9/tracrARN/crARN escinda una secuencia diana de ADN de cadena doble, la secuencia de ADN diana está adyacente a un PAM equivalente. Modificando un crARN para que tenga una secuencia espaciadora apropiada, el complejo puede dirigirse para escindirlo en el locus de interés, por ejemplo, un locus en el que se desee la modificación de secuencia.

Ran, F.A., et al, *Nature* 520(7546):186-191 (2015), incluyendo todos los datos extendidos, presentan las secuencias crARN/tracrARN y las estructuras secundarias de ocho sistemas CRISPR-Cas Tipo II (véase, la Figura 1 de los datos extendidos de Ran, F.A., et al). Las estructuras de tracrARN previstas se basaban en el modelo de plegado de ARN por Generación rigurosa (Zuker, M., *Nucleic Acids Research* 31:3406-3415 (2003)). Además, Fonfara, et al, *Nucleic Acids Research* 42(4):2577-2590 (2014), incluyendo todos los datos suplementarios (en particular la Figura Suplementaria S11) presentan secuencias y estructuras secundarias de crARN/tracrARN de ocho sistemas CRISPR-

Cas Tipo II. Las estructuras secundarias del dúplex de ARN se previeron utilizando el RNAcofold del paquete Vienna RNA (Bernhart, S.H., et al, Algorithms for Molecular Biology 1(1):3 (2006); Hofacker, I.L., et al, Journal of Molecular Biology 319:1059-1066 (2002)) y RNAhybrid (bibiserv.techfak.uni-bielefeld.de/rmahybrid/). Las previsiones de la estructura se visualizaron utilizando VARNA (Darty, K., et al, Bioinformatics 25:1974-1975 (2009)). Fonfara, et al, demostraron que el complejo crARN/tracrARN de *Campylobacter jejuni* no tiene la región de protuberancia; sin embargo, mantiene una estructura en tallo localizada 3' del espaciador que continúa en dirección 3' con otra estructura en tallo.

Los sistemas CRISPR-Cas de Tipo V de origen natural, a diferencia de los sistemas CRISPR Car Tipo II, no necesitan un tracrARN para la maduración de crARN y la escisión de una secuencia de ácido nucleico diana. En una estructura típica de un crARN del CRISPR del Tipo V, la secuencia que se une al ADN diana está corriente abajo de una estructura secundaria específica (es decir, una estructura tallo-lazo) que interactúa con la proteína Cpf1. Las bases 5' del tallo lazo adoptan una estructura de pseudo-nudo que estabiliza adicionalmente la estructura tallo-lazo con un emparejamiento de bases de Watson-Crick no canónico, interacción triple, y emparejamiento de bases de Hoogsteen (véase Yamano, T., et al, Cell 165(4):949-962 (2016)).

El espaciador de los sistemas CRISPR-Cas de Clase 2 se puede hibridar a una secuencia de ácido nucleico diana que se localiza 5' o 3' de un PAM, dependiendo de la proteína Cas que se va a utilizar. Un PAM puede variar dependiendo del polipéptido Cas que se va a utilizar. Por ejemplo, si se utiliza la Cas) de *S. pyogenes*, el PAM puede ser una secuencia en la secuencia del ácido nucleico diana que comprende la secuencia 5'-NRR-3', en el que R puede ser cualquiera de A o G, en el que N es cualquier nucleótido, y N está inmediatamente 3' de la secuencia de ácido nucleico diana dirigida por la secuencia de unión del ácido nucleico diana. Una proteína Cas puede modificarse de manera que un PAM puede ser diferente en comparación con un PAM de una proteína Cas sin modificar. Por ejemplo, si se utiliza la Cas9 de *S. pyogenes*, la proteína Cas9 puede modificarse de manera que el PAM no comprenda la secuencia 5'-NRR-3', sino que comprende en vez de esta la secuencia 5'-NNR-3', en la que R puede ser cualquiera de A o G, en la que N es cualquier nucleótido, y N está inmediatamente 3' de la secuencia de ácido nucleico diana dirigida por la secuencia de ácido nucleico diana.

Otras proteínas Cas reconocen otros PAM y un experto en la técnica es capaz de determinar el PAM para cualquier proteína Cas particular. Por ejemplo, Cpf1 tiene un sitio PAM rico en timina que se dirige, por ejemplo, a una secuencia TTTN (véase Fagerlund, R., et al, Genome Biology 16:251 (2015)).

La endonucleasa Cas9 guiada por ARN se ha utilizado ampliamente para la edición de genoma programable en varios organismos y sistemas modelo (véase, por ejemplo, Jinek M., et al., Science 337:816-821 (2012); Jinek M., et al, eLife 2: e00471. doi: 10.7554/eLife 00471 (2013); Solicitud de Patente Publicada de EE. UU N.º 2014-0068797, publicada el 6 de marzo de 2014).

La modificación del genoma incluye la alteración del genoma por eliminación, inserción, mutación o sustitución de secuencias de ácido nucleico específicas. La alteración puede ser específica del gen o de la localización. La modificación del genoma puede utilizar nucleasas dirigidas al sitio, tales como las proteínas Cas y sus polinucleótidos equivalentes, para cortar el ADN, generando de esta manera un sitio para la alteración. En ciertos casos, la escisión puede introducir una rotura de cadena doble (DSB) en la secuencia de ADN diana. Las DSB se pueden reparar, por ejemplo, mediante una unión de extremos no homóloga (NHEJ), unión de extremos mediada por microhomología (MMEJ), o reparación dirigida por homología (HDR). La HDR se basa en la presencia de una matriz para reparar. En algunos ejemplos de modificación del genoma se puede insertar un polinucleótido donante o parte del mismo en la rotura.

### Sumario de la invención

La invención se define en las reivindicaciones adjuntas.

La presente invención se refiere en general a ácidos nucleicos modificados de tipo cruzado que se dirigen a ácidos nucleicos.

En un aspecto, la presente invención se refiere a un ácido nucleico modificado de tipo cruzado con CRISPR de Clase 2 que se dirige a un ácido nucleico ("NATNA de tipo cruzado con CRISPR de Clase 2"), que comprende: un ácido nucleico asociado con Cpf1 que se dirige a un ácido nucleico ("Cpf1-NATNA"); un primer ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico, que tiene un extremo 5' y un extremo 3', que comprende un elemento espaciador ("primer Cas9-NATNA"); y un segundo ácido nucleico asociado con Cas9 dirigido a un ácido nucleico, que tiene un extremo 5' y un extremo 3', que comprende un elemento tracr ("segundo Cas9-NATNA"); en el que el primer Cas9-NATNA o el segundo Cas9-NATNA está conectado con el Cpf1-NATNA. En una realización, el primer Cas9-NATNA o el segundo Cas9-NATNA se conecta no covalentemente con el Cpf1-NATNA, por ejemplo, mediante, unión de hidrógeno de pares de bases en el extremo 5' o el extremo 3' del primer Cas9-NATNA o mediante unión de hidrógeno de pares de bases en el extremo 5' o 3' del segundo Cas9-NATNA. En una realización alternativa, el primer Cas9-NATNA o el segundo Cas9-NATNA se conecta covalentemente con el Cpf1-NATNA, por ejemplo, el Cpf1-NATNA se conecta covalentemente con el extremo 5' o el extremo 3' del primer Cas9-NATNA o con el extremo 5' o el extremo

3' del segundo Cas9-NATNA.

En algunas realizaciones, el CpfI-NATNA de los NATNA tipo cruzado del CRISPR Clase 2 es capaz de formar un primer complejo con la proteína CpfI ("complejo Cas9-CpfI-NATNA/proteína CpfI"), y el primer Cas9-NATNA de los NATNA tipo cruzado con CRISPR Clase 2 y el segundo Cas9-NATNA del NATNA tipo cruzado con CRISPR Clase 2 son capaces de formar un segundo complejo con una proteína Cas9 ("complejo Cas9-CpfI-NATNA/proteína Cas9"). Si el CpfI-NATNA forma el primer complejo con la proteína CpfI ("complejo Cas9-CpfI-NATNA/proteína CpfI"), el primer complejo es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble complementaria al elemento espaciador CpfI. Además, si el primer Cas9-NATNA y el segundo Cas9-NATNA forman el segundo complejo con la proteína Cas9 ("complejo Cas9-NATNA/proteína Cas9"), el segundo complejo es capaz de unirse a una segunda secuencia del ácido nucleico diana de cadena doble complementaria al primer elemento espaciador de Cas9-NATNA. En otras realizaciones, si se forma el primer complejo y se forma el segundo complejo, dando como resultado un complejo Cas9-CpfI-NATNA/proteínas Cas9&CpfI, el Cas9-CpfI-NATNA/Cas9&CpfI es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble complementaria al elemento espaciador de CpfI y a una segunda secuencia de ácido nucleico diana complementaria al primer elemento espaciador de Cas9-NATNA.

En realizaciones adicionales, uno o más de entre el CpfI-NATNA, el primer Cas9-NATNA, y el segundo Cas9-NATNA pueden comprender adicionalmente un elemento enlazador de la secuencia de nucleótidos conectado covalentemente con el extremo 5' o el extremo 3' de uno o más de entre el CpfI-NATNA, el primer Cas9-NATNA y el segundo Cas9-NATNA. Además, un primer elemento enlazador de la secuencia de nucleótidos y un segundo elemento enlazador de la secuencia de nucleótidos pueden conectarse covalentemente como ambos extremos 5' y 3' de uno o más de entre el CpfI-NATNA, el primer Cas9-NATNA y el segundo Cas9-NATNA.

En realizaciones de la presente invención también incluyen el NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 3' del primer Cas9-NATNA está conectado covalentemente mediante un elemento en bucle con el extremo 3' del segundo Cas9-NATNA. La conexión covalente da como resultado un ácido nucleico asociado con Cas9 sencillo que se dirige a ácidos nucleicos ("Cas9-NATNA-sencillo"), que tiene un extremo 5' y un extremo 3'. El Cas9-NATNA-sencillo comprende el primer Cas9-NATNA y el segundo Cas9-NATNA. En algunas realizaciones el Cas9-NATNA-sencillo está conectado no covalentemente con el CpfI-NATNA mediante unión de hidrógeno de pares de bases en el extremo 5' o el extremo 3'. Además, un primer CpfI-NATNA y un segundo CpfI-NATNA pueden estar conectados no covalentemente con ambos extremos 5' y 3' del Cas9-NATNA-sencillo. En otras realizaciones, el Cas9-NATNA-sencillo está conectado covalentemente con el CpfI-NATNA mediante unión de hidrógeno de pares de bases en el extremo 5' o el extremo 3'. Además, un primer CpfI-NATNA y un segundo CpfI-NATNA pueden estar conectados covalentemente con ambos extremos 5' y 3' del Cas9-NATNA-sencillo.

En realizaciones adicionales, uno o ambos del CpfI-NATNA y el Cas9-NATNA-sencillo pueden comprender adicionalmente un elemento enlazador de la secuencia de nucleótidos conectado covalentemente con el extremo 5' o el extremo 3'. Además, un primer elemento enlazador de la secuencia de nucleótidos y un segundo elemento enlazador de la secuencia de nucleótidos pueden conectarse covalentemente con ambos extremos 5' y 3' de uno o ambos del CpfI-NATNA y el Cas9-NATNA-sencillo.

En algunas realizaciones, uno o más de entre el CpfI-NATNA, el primer Cas9-NATNA y el segundo Cas9-NATNA (incluyendo el Cas9-NATNA-sencillo que comprende el primer Cas9-NATNA y el segundo Cas9-NATNA) del NATNA de tipo cruzado con CRISPR de Clase 2 modificado puede comprender un ARN, ADN, o combinaciones de ARN y ADN.

En un aspecto adicional, la presente invención incluye una composición ácido nucleico/proteína que comprende un NATNA tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cas9, y una proteína CpfI. En algunas realizaciones, el NATNA tipo cruzado con CRISPR de Clase 2 modificado está formando un complejo con la proteína Cas9 y la proteína CpfI. En realizaciones adicionales, una o ambas de entre la proteína CpfI y la proteína Cas9 pueden ser enzimáticamente inactivas. Cuando la proteína CpfI es enzimáticamente inactiva (proteína dCpfI) o la proteína Cas9 es enzimáticamente inactiva (proteína dCas9), la composición ácido nucleico/proteína puede comprender adicionalmente un polinucleótido donante conectado no covalentemente con la proteína dCpfI o la proteína dCas9.

En otro aspecto, la presente invención se refiere a un vector de expresión, que comprende una o más secuencias de ácido nucleico que codifican un NATNA tipo cruzado con CRISPR de Clase 2 modificado.

En otro aspecto más, la presente invención incluye una célula recombinante, que comprende una o más secuencias de ácido nucleico que codifican un NATNA tipo cruzado con CRISPR de Clase 2 modificado.

Un aspecto adicional de la presente invención incluye un kit que comprende un NATNA de tipo cruzado con CRISPR de Clase 2 modificado y un tampón, o una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado y un tampón. Los kits pueden comprender adicionalmente una proteína Cas9, una proteína CpfI, o ambas, una proteína Cas9 y una proteína CpfI. Además, los kits pueden comprender una o más secuencias de ácido nucleico que codifican una proteína Cas9, una proteína CpfI, o ambas, una proteína Cas9 y una proteína CpfI.

Aspectos adicionales de la presente invención incluyen métodos de utilización de un NATNA tipo cruzado con CRISPR de Clase 2 modificado como se describe en el presente documento. Un método es un método de unión al ADN, en el que el método no comprende un método para el tratamiento del cuerpo humano o animal. El método comprende poner en contacto de un a primera secuencia de ADN diana de un ADN con una segunda secuencia de ADN diana de un ADN con una composición de ácido nucleico/proteína que comprende un NATNA tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cas9, y una proteína Cpf1; facilitando de esta manera la unión de la composición de ácido nucleico/proteína la primera secuencia de ADN diana de un ADN y la segunda secuencia de ADN diana de un ADN. El elemento espaciador Cpf1-NATNA del NATNA tipo cruzado con CRISPR de Clase 2 es complementario de la primera secuencia de ADN diana, y el espaciador Cas9-NATNA del NATNA tipo cruzado con CRISPR de Clase 2 es complementario de la segunda secuencia del ADN diana.

Otro método de la presente invención es un método de corte de un ADN, en el que el método no comprende un proceso para la modificación de identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El método comprende poner en contacto de una primera secuencia de ADN diana en un ADN y una segunda secuencia de ADN diana en un ADN con una composición de ácido nucleico/proteína que comprende un NATNA tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cas9, y una proteína Cpf1; facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia de ADN diana y la segunda secuencia de ADN diana que da como resultado el corte de la primera secuencia de ADN diana y la segunda secuencia de ADN diana. El elemento espaciador Cpf1-NATNA del NATNA de tipo cruzado con CRISPR de Clase 2 es complementario de la primera secuencia de ADN diana, y el espaciador Cas9-NATNA del NATNA tipo cruzado con CRISPR de Clase 2 es complementario de la segunda secuencia de ADN diana. Además, la proteína Cpf1 de la composición de ácido nucleico/proteína unida es capaz de cortar la primera secuencia de ADN diana y la proteína Cas9 de la composición de ácido nucleico/proteína unida es capaz de cortar la segunda secuencia de ADN diana.

Estos aspectos y otras realizaciones de la presente invención que utilizan los sistemas CRISPR-Cas de Clase 2 modificado de la presente invención serán fácilmente evidentes para los expertos en la técnica en vista de la divulgación del presente documento.

### Breve descripción de las figuras

Las figuras no se han hecho proporcionalmente, ni están a escala. Las localizaciones de los indicadores son aproximadas.

La FIG. 1A y FIG. 1B presentan ejemplos ilustrativos de ARN guía asociados a CRISPR de Clase 2 Tipo II de doble guía.

La FIG. 2 presenta un ejemplo ilustrativo de un ARN guía asociado a CRISPR de Clase 2 Tipo II guía sencillo.

La FIG. 3 presenta un ejemplo ilustrativo de un ARN guía crARN de Clase 2 Tipo V.

Las FIG. 4, FIG. 5A, FIG. 5B, FIG. 6, FIG. 7, FIG. 8, FIG. 9 y FIG. 10 presenta ejemplos de ácido nucleico Cas9-Cpf1 modificado que se dirige a ácidos nucleicos de la presente invención.

Las Fig. 11A, FIG. 11B, FIG. 11C, FIG. 11D ilustran la escisión de una secuencia de ácido nucleico diana utilizando un ácido nucleico Cas9-Cpf1 modificado que se dirige a un ácido nucleico y proteínas Cas9 y Cpf1 que tienen ambas endonucleasas activas.

Las FIG. 12A, FIG. 12B, y FIG. 12C ilustran la utilización de un ácido nucleico Cas9-Cpf1 modificado que se dirige a un ácido nucleico y proteínas Cas9 y Cpf1, en las que los dominios endonucleasa de Cas9 son activos y el dominio endonucleasa de Cpf1 está inactivo, para poner un polinucleótido donante en proximidad de un DSB en una secuencia de ácido nucleico diana.

Las FIG. 13A, FIG. 13B y FIG. 13C ilustran la utilización de un ácido nucleico Cas9-Cpf1 modificado que se dirige a un ácido nucleico y proteínas Cas9 y Cpf1, en las que los dominios endonucleasa de Cas9 están inactivos y el dominio endonucleasa de Cpf1 está activo, para poner un polinucleótido donante en proximidad de una rotura de cadena doble en una secuencia de ácido nucleico diana.

Las FIG. 14A hasta la FIG. 14E representan el Cpf1-crARN. La FIG. 14A ilustra un Cpf1-crARN de tipo silvestre. La FIG. 14B ilustra un Cpf1 pre-crARN de tipo silvestre. La FIG. 14C, FIG. 14D y FIG. 14E ilustra Cpf1-crARN modificados.

La FIG. 15 presenta los resultados de los ensayos de escisión bioquímica de Cpf1 de F. novicida utilizando distintos Cpf1 crARN modificados.

### Descripción detallada de la invención

La invención se define en las reivindicaciones adjuntas.

Se tiene que entender que la terminología utilizada en el presente documento tiene el fin de describir realizaciones particulares solamente, y no pretende ser limitante. Como se utiliza en la presente memoria descriptiva y las reivindicaciones adjuntas, las formas singulares “un”, “una” y “el” incluyen los referentes en plural a menos de que el contexto dicte claramente otra cosa. Por lo tanto, por ejemplo, la referencia a “un polipéptido” incluye uno o más

polinucleótidos, y la referencia a “un vector” incluye uno o más vectores.

A menos de que se defina otra cosa, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que entiende comúnmente un experto habituado en la técnica a la que pertenece la invención. Aunque pueden ser útiles en la presente invención otros métodos y materiales similares, o equivalentes, a los que se describen en el presente documento, se describen en el presente documento los materiales y métodos preferidos.

En vista de las enseñanzas de la presente memoria descriptiva, un experto habituado en la técnica puede emplear técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica, y polinucleótidos recombinantes, como se enseña, por ejemplo, en los siguientes textos convencionales: Anticuerpos: A Laboratory Manual, Segunda edición, E. A. Greenfield, Cold Spring Harbor Laboratory Press, ISBN 978-1-936113-81-1 (2014); Cultivo de células animales: A Manual of Basic Technique and Specialized Applications, 6ª Edición, R. I. Freshney, Wiley-Blackwell, ISBN 978-0-470-52812-9 (2010); Transgenic Animal Technology, tercera Edición: A Laboratory Handbook, C. A. Pinkert, Elsevier, ISBN 978-0124104907 (2014); The Laboratory Mouse, segunda Edición, H. Hedrich, Academic Press, ISBN 978-0123820082 (2012); Manipulating the Mouse Embryo: A Laboratory Manual, R. Behringer, et al, Cold Spring Harbor Laboratory Press, ISBN 978-1936113019 (2013); PCR 2: A Practical Approach, M. J. McPherson, et al, IRL Press, ISBN 978-0199634248 (1995); Methods in Molecular Biology (Series), J.M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, D. C. Rio, et al, Cold Spring Harbor Laboratory Press, ISBN 978-0879698911 (2010); Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (Cuarta Edición), M. R. Green, et al, Cold Spring Harbor Laboratory Press, ISBN 978-1605500560 (2012); Bioconjugate Techniques, tercera Edición, G. T. Hermanson, Academic Press, ISBN 978-0123822390 (2013); Methods in Plant Biochemistry and Molecular Biology, W. V. Dashek, CRC Press, ISBN 978-0849394805 (1997); Plant Cell Culture Protocols (Methods in Molecular Biology), V. M. Loyola-Vargas, et al, Humana Press, ISBN 978-1617798177 (2012); Plant Transformation Technologies, C. N. Stewart, et al, Wiley-Blackwell, ISBN 978-0813821955 (2011); Recombinant Proteins from Plants (Methods in Biotechnology), C. Cunningham, et al, Humana Press, ISBN 978-1617370212 (2010); Plant Genomics: Methods and Protocols (Methods in Molecular Biology), D. J. Somers, et al, Humana Press, ISBN 978-1588299970 (2009); Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, R. Keshavachandran, et al, Orient Blackswan, ISBN 978-8173716164 (2008).

Las repeticiones cortas palindrómicas interespaciadas regularmente agrupadas (CRISPR) y las proteínas asociadas a CRISPR relacionadas (proteínas Cas) constituyen los sistemas CRISPR-Cas (véase, Barrangou, R., et al, Science 315:1709-1712 (2007)).

Como se utiliza en el presente documento, “proteína Cas” y “proteína CRISPR-Cas” se refiere a proteínas asociadas a CRISPR (Cas) que incluyen, pero no se limitan a proteínas asociadas a CRISPR de Clase 1 Tipo 1, proteínas asociadas a CRISPR de Clase 1 Tipo III, y proteínas asociadas a CRISPR de Clase 1 Tipo IV, proteínas asociadas a CRISPR de Clase 2 Tipo II, proteínas asociadas a CRISPR de Clase 2 Tipo V y proteínas asociadas a CRISPR de Clase 2 Tipo VI. Las proteínas Cas de Clase 2 incluyen proteínas Cas9, proteínas tipo Cas9 codificadas por ortólogos Cas9, proteínas sintéticas tipo Cas9, proteínas Cpf1, proteínas codificadas por ortólogos Cpf1, proteínas sintéticas tipo Cpf1, proteínas C2c1, proteínas C2c2, proteínas C2c3, y variantes y modificaciones de las mismas. En algunas realizaciones las proteínas son proteínas asociadas a CRISPR de Clase 2, por ejemplo, una o más proteínas asociadas a CRISPR de Clase 2 Tipo II, tales como Cas 9, una o más proteínas asociadas a CRISPR de Clase 2 Tipo V, tal como Cpf1 y una o más proteínas asociadas a CRISPR de Clase 2 tipo VI, tales como C2c2. En realizaciones preferidas, las proteínas Cas son una o más proteínas asociadas a CRISPR de Clase 2 Tipo II, tales como Cas 9, y una o más proteínas asociadas a CRISPR de Clase 2 Tipo V, tales como Cpf1. Normalmente, para su uso en los aspectos de la presente invención, una proteína Cas es capaz de interactuar con uno o más polinucleótidos equivalentes (más normalmente ARN) para formar un complejo nucleoproteico (más normalmente, un complejo ribonucleoproteico).

“Proteína Cas9” como se utiliza en el presente documento, se refiere a una proteína Cas9 de tipo silvestre derivada de sistemas CRISPR de Clase 2 Tipo II-Cas9, modificaciones de proteínas Cas9, variantes de proteínas Cas9, ortólogos de Cas9, y combinaciones de los mismos. Las proteínas Cas9 incluyen, pero no se limitan a Cas9 de *Streptococcus pyogenes* (UniProtKB - Q99ZW2 (CAS9STRP1)), *Streptococcus thermophilus* (UniProtKB - G3ECR1 (CAS9STRTR)), y *Staphylococcus aureus* (UniProtKB - J7RUA5 (CAS9STAAU)). Los homólogos de Cas9 se pueden identificar utilizando métodos de búsqueda de similitud de secuencia conocidos por el experto en la técnica. “dCas9”, como se utiliza en el presente documento, se refiere a variantes de la proteína Cas9 que son proteínas Cas9 con nucleasa desactivada, también llamada “proteína Cas9 catalíticamente inactiva”, “Cas9 enzimáticamente inactiva”, “Cas9 catalíticamente muerta” o “Cas9 muerta”. Dichas moléculas carecen de toda o una parte de la actividad endonucleasa y por lo tanto se pueden utilizar para regular genes de una manera guiada por ARN (véase Jinek M., et al, Science 337:816-821 (2012)). Esto se consigue introduciendo mutaciones que inactivan la función nucleasa de Cas9 y se consigue normalmente mutando ambos restos catalíticos (D10A en el dominio RuvC-1, y H840A en el dominio HNH, numerados respecto a la Cas9 de *S. pyogenes*). Se entiende que la mutación de otros restos para reducir la actividad de cualquiera o ambos dominios de nucleasa también se pueden llevar a cabo por un experto en la técnica. La dCas9 resultante es incapaz de escindir un ADN de cadena doble, pero mantiene la capacidad para formar un complejo con un ácido nucleico guía y unirse a una secuencia de ADN diana. La Cas9 mutante doble con

cambios en las posiciones de aminoácidos D10A y H840A inactiva completamente ambas actividades nucleasa y nickasa. La especificidad de direccionamiento se determina por emparejamiento de bases complementarias del ARN guía (normalmente un ARN guía sencillo) al locus genómico y el PAM. Cas9 es una proteína de firma característica de los sistemas con CRISPR de Clase 2 Tipo II.

5 “Proteína Cpfl”, como se utiliza en el presente documento, se refiere a una proteína Cpfl de tipo silvestre derivada de sistemas CRISPR de Clase 2 Tipo V-Cpfl, modificaciones de proteínas Cpfl, variantes de proteínas Cpfl, ortólogos de Cpfl, y combinaciones de los mismos. “dCpfl” como se utiliza en el presente documento, se refiere a variantes de la proteína Cpfl que son proteínas Cpfl con nucleasa desactivada, también llamada “proteína Cpfl catalíticamente inactiva”, o “Cpfl enzimáticamente inactiva”. Las proteínas Cpfl incluyen, pero no se limitan a, *Francisella novicida* (UniProtKB - A0Q7Q2 (CPFLFRATN)), *Lachnospiraceae bacterium* (UniProtKB - A0A182DWE3 (A0A182DWE39FIRM)), y *Acidaminococcus sp.* (UniProtKB - U2UMQ6 (CPFLACISB)). Cpfl es una proteína firma característica de los sistemas CRISPR de Clase 2 Tipo V. Los homólogos de Cpfl se pueden identificar utilizando métodos de búsqueda de similitud de secuencia conocidos por el experto en la técnica.

15 “Proteína argonauta”, como se utiliza en el presente documento se refiere a una proteína Argonauta de tipo silvestre, modificaciones de proteínas Argonauta, variantes de proteínas Argonauta, ortólogos de Argonauta, y combinaciones de los mismos (véase, por ejemplo, Hall, T., Structure 13:1403-1408 (2005); Hock, J., et al, Genome Biology 9(2): 210-210.8 (2008); Swarts, D., et al, Nature 507(7491):258-261 (2014); Swarts, D., et al, Nature Structural & Molecular Biology (9):743-753 (2014); Hur, J.K., et al, Trends in Biochemical Sciences 39(6):257-259 (2014)). “dArgonauta”, como se utiliza en el presente documento, se refiere a variantes de proteínas Argonauta que son proteínas Argonauta con nucleasa desactivada, también llamada “proteína Argonauta catalíticamente inactiva”, o “Argonauta enzimáticamente inactiva”. “Proteína Argonauta se refiere a una proteína de una familia de proteínas normalmente definida por la presencia de un dominio PIWI y/o un dominio PAZ (PIWI-Argonauta-Zwille). Una proteína Argonauta (por ejemplo, una Argonauta eucariota o Argonauta procariota) normalmente es capaz de interactuar con una guía Argonauta (es decir, un ácido nucleico que se dirige a ácidos nucleicos; Ago-NATNA) para formar un complejo. El Ago-NATNA comprende una secuencia de ácido nucleico de unión a la diana. El complejo es capaz de unirse de manera dirigida al sitio a una secuencia de ácido nucleico diana. El complejo se dirige a la secuencia de ácido nucleico diana por la secuencia de ácido nucleico de unión a la diana del Ago-NATNA. Las secuencias de ácido nucleico diana a las que se une el complejo proteína Argonauta/Ago-NATNA pueden ser, por ejemplo, ARN, ADN o híbridos de ARN/ADN. Los homólogos de proteína Argonauta se pueden identificar utilizando métodos de búsqueda de similitud de secuencia conocidos por el experto en la técnica.

35 “Ácido nucleico que se dirige a ácidos nucleicos” (NATNA) significa uno o más polinucleótidos que guían una proteína, tal como una proteína Argonauta o más preferentemente una proteína Cas (por ejemplo, una proteína Cas9, una proteína dCas9, una proteína Cpfl, o una proteína dCpfl) para unirse preferentemente a una secuencia de ácido nucleico diana, normalmente, una secuencia diana de ácido nucleico de cadena doble. Por ejemplo, unas realizaciones de la presente invención incluyen, pero no se limitan a, NATNA modificados que comprenden un ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico y un ácido nucleico asociado con Cpfl que se dirige a un ácido nucleico. Ejemplos de estos NATNA incluyen, pero no se limitan a los siguientes: un ácido nucleico asociado con Cas que se dirige a un ácido nucleico que comprende un elemento espaciador (“primer Cas9-NATNA”; por ejemplo, FIG. 1A, 101) y un ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico que comprende un elemento tracr (“segundo Cas9-NATNA”, por ejemplo, FIG. 1A, 102); un “Cas9-NATNA-sencillo” (véase, por ejemplo, la FIG. 2) que comprende un primer Cas9-NATNA y un segundo Cas9-NATNA; y un ácido nucleico asociado con Cpfl que se dirige a un ácido nucleico que comprende un elemento espaciador (“Cpfl-NATNA”) (véase, por ejemplo, la FIG. 3). Ejemplos de Cas9- Cpfl-NATNA de un aspecto de la presente invención incluyen, pero no se limitan a, los que se ilustran en la FIG. 3, FIG. 4, FIG. 5A, FIG. 5B, FIG. 6, FIG. 7, FIG. 8, FIG. 9, y FIG. 10. Los NATNA pueden comprender bases de ribonucleótidos (por ejemplo, ARN), bases de desoxirribonucleótidos (por ejemplo, ADN), combinaciones de bases de ribonucleótidos y desoxirribonucleótidos (por ejemplo, ARN/ADN), nucleótidos, análogos de nucleótidos, nucleótidos modificados, y similares, así como restos de estructura o uniones modificados sintéticos, de origen natural y de origen no natural, por ejemplo, como se describe en el presente documento.

Como se utiliza en el presente documento, “ARN doble guía” y “Cas9-ARN doble guía” se refiere normalmente a un sistema de ARN de dos componentes para un componente polinucleotídico capaz de asociarse con una proteína Cas9 equivalente. La FIG. 1A y la Fig. 1B presentan ejemplos ilustrativos de ARN asociados a Cas9-CRISPR de Clase 2 Tipo II. La FIG. 1A muestra un sistema de dos componentes de ARN con CRISPR-Cas9 Tipo II que comprende un Cas9-crARN (FIG. 1A, 101) y un Cas9-tracrARN (FIG. 1A, 102). La FIG. 1B ilustra la formación de enlaces de hidrógeno de pares de bases entre el Cas9-crARN y el Cas9-tracrARN para formar una estructura secundaria (véase la Solicitud de Patente Publicada N.º 2014-0068797, publicada el 6 de marzo de 2017; véase también Jinek M., et al, Science 337:816-21 (2012)). La FIG. 1B presenta una panorámica y una nomenclatura para los elementos estructurales secundarios del Cas9-crARN y Cas9-tracrARN de la Cas9 de *S. pyogenes* incluyendo lo siguiente: un elemento espaciador (FIG. 1B, 103); un primer elemento tallo que comprende un elemento tallo inferior (FIG. 1B, 104), un elemento protuberante que comprende nucleótidos no emparejados (FIG. 1B, 105), y un elemento tallo superior (FIG. 1B, 106); un elemento de nexa (FIG. 1B, 107); un primer elemento ahorquillado 3’ (FIG. 1B, 108); y un segundo elemento horquillado 3’ (FIG. 1B, 109). Un Cas-ARN doble guía es capaz de formar un complejo nucleoproteico con una proteína Cas9 equivalente, en la que el complejo es capaz de dirigirse a una secuencia de ácido nucleico diana

complementaria a la secuencia espaciadora. Las modificaciones de Cas9 doble guía se conocen en la técnica, incluyendo, la eliminación de uno o más elementos ahorquillados 3' (FIG. 1B, 108, 109) y modificaciones en el tallo superior, protuberancia, y tallo inferior (FIG. 1B 106, 105, 104, respectivamente) (véase, por ejemplo, la Publicación de Patente de EE. UU. N.º 2014-0315985, publicada el 23 de octubre de 2014; Publicación de Patente de EE. UU. N.º 2015-0376586, publicada el 31 de diciembre de 2015).

Como se utiliza en el presente documento, "ARN guía sencillo" (sgARN) y "Cas9-sgARN" se refiere normalmente a un sistema de ARN de un componente para un componente polinucleotídico capaz de asociarse con una proteína Cas9 equivalente. La FIG. 2 muestra un ejemplo de un sgARN asociado con CRISPR de Clase 2 Tipo II-Cas9. La figura ilustra un ARN guía sencillo Cas9 (Cas9-sgARN) en el que el Cas9-crARN está unido covalentemente al Cas9-tracrARN y forma una estructura secundaria de ARN polinucleotídico mediante unión de hidrógeno de pares de bases (véase, por ejemplo, la Solicitud de Patente Publicada de EE. UU. N.º 2014-0068797, publicada el 6 de marzo de 2014). La FIG. 2 presenta una panorámica y nomenclatura para los elementos estructurales secundarios de un Cas9-sgARN para *S. pyogenes* que incluye lo siguiente: un elemento espaciador (FIG. 2, 201); un primer elemento tallo que comprende un elemento tallo inferior (DIF. 2, 202), un elemento protuberante que comprende nucleótidos no emparejados (FIG. 2, 205), y un elemento de tallo superior (FIG. 2, 203); un elemento de bucle (FIG. 2, 204) que comprende nucleótidos no emparejados; un elemento de nexa (FIG. 2, 206); un primer elemento ahorquillado 3' (FIG. 2, 207); y un segundo elemento ahorquillado 3' (FIG. 2, 208), (véase, por ejemplo, las Figuras 1 y 3 de Briner, A. E., et al., Molecular Cell 56(2):333-339 (2014)). Un Cas9-sgARN es capaz de formar un complejo nucleoproteico con una proteína Cas9 equivalente, en el que el complejo es capaz de dirigirse a una secuencia de ácido nucleico complementaria a la secuencia espaciadora. Las modificaciones de Cas9 guías únicas se conocen en la técnica, incluyendo la eliminación de uno o más elementos ahorquillados 3' (FIG. 2, 207, 208) y las modificaciones del tallo superior, protuberancia y tallo inferior (FIG. 2, 203, 205, 202, respectivamente) (véase, por ejemplo, la Publicación de Patente de EE. UU. N.º 2014-0315985, publicada el 23 de octubre de 2014, Publicación de Patente de EE. UU. N.º 2015-0376586, publicado el 31 de diciembre de 2015).

"crARN guía" y "Cpfl-ARN", como se utiliza en el presente documento, se refiere normalmente a un sistema de un componente ARN para un componente polinucleotídico capaz de asociarse con una proteína Cpfl equivalente. La FIG. 3 presenta un ejemplo de un ARN asociado con CRISPR Tipo V-Cpfl (Cpfl-crARN) (véase, por ejemplo, Zetsche, B., et al, Cell 163:1-13 (2015)). La FIG. 3 presenta una panorámica y la nomenclatura para elementos estructurales secundarios de un Cpfl-crARN de la siguiente manera: un elemento tallo-lazo (FIG. 3, 301) y un elemento espaciador (FIG. 3, 302). El elemento tallo-lazo comprende, en dirección 5' a 3', una secuencia 1 de tallo del ARN-Cpfl (FIG. 3, 303), un elemento bucle (FIG. 3, 304), y una secuencia 2 del tallo de ARN-Cpfl (FIG. 3, 305), en el que la secuencia 1 del tallo de ARN-Cpfl y la secuencia 2 del tallo de ARN-Cpfl complementaria forman un dúplex. Un crARN guía es capaz de formar un complejo nucleoproteico con una proteína Cpfl equivalente, en el que el complejo es capaz de dirigirse a una secuencia de ácido nucleico diana complementaria a la secuencia espaciadora.

Como se utiliza en el presente documento, un "ácido nucleico tipo cruzado que se dirige a un ácido nucleico" y un "NATNA tipo cruzado" se utilizan de manera intercambiable para referirse a un primer NATNA conectado con un segundo NATNA. Por ejemplo, un NATNA tipo cruzado con CRISPR de Clase 2 se refiere a uno o más polinucleótidos que comprenden normalmente un NATNA con CRISPR de Clase 2 Tipo V (por ejemplo, la FIG. 3) conectado con un NATNA con CRISPR de Clase 2 Tipo II (por ejemplo, FIG. 1A, FIG. 1B, FIG. 2). Los tipos de conexiones entre el primer NATNA y el segundo NATNA para formar el NATNA de tipo cruzado incluyen, por ejemplo, el enlace covalente (por ejemplo, la FIG. 4), el enlace de hidrógeno (por ejemplo, FIG. 5A, FIG. 5B), emparejamiento de restos por unión ligando/ligando, y/o entrecruzamiento. Los enlaces covalentes incluyen, pero no se limitan a enlaces fosfodiéster. Un NATNA de tipo cruzado es capaz de formar un complejo con una proteína que es capaz de formar un complejo con el primer NATNA y una segunda proteína que es capaz de formar un complejo con el segundo NATNA. Por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 es capaz de formar un complejo con una proteína Cas de Clase 2 Tipo V (por ejemplo, una proteína, Cpfl) que es capaz de formar un complejo con el NATNA con CRISPR de Clase 2 Tipo V y la proteína Cas de Clase 2 Tipo II (por ejemplo, una proteína Cas9) que es capaz de formar un complejo con el NATNA con CRISPR de Clase 2 Tipo II. En esta realización, un complejo formado entre el NATNA-Cpfl y una proteína Cpfl es capaz de unirse a una secuencia de ácido nucleico diana de cadena doble. También a un complejo formado entre el Cas9-NATNA (por ejemplo, un ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico que comprende un elemento espaciador, "primer Cas9-NATNA", y un segundo ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico que comprende un elemento tracr, "segundo Cas9-NATNA", o un Cas9-NATNA sencillo) y una proteína Cas9 es capaz de unirse una segunda secuencia de ácido nucleico diana de cadena doble (por ejemplo, FIG. 11B). En otra realización, un NATNA de tipo cruzado comprende un Ago-NATNA y un NATNA de Clase 2 Tipo V o un NATNA de Clase 2 Tipo II.

Como se utiliza en el presente documento, un "entrecruzamiento" es un enlace que une una cadena polimérica (por ejemplo, un polinucleótido o polipéptido) a otro. Dichos enlaces pueden ser enlaces covalentes o enlaces iónicos. Por ejemplo, un polinucleótido se puede unir a otro polinucleótido mediante entrecruzamiento de los polinucleótidos.

Como se utiliza en el presente documento, el término "equivalente" normalmente se refiere a una proteína Cas y uno o más polinucleótidos Cas que son capaces de formar un complejo nucleoproteico capaz de la unión dirigida al sitio a una secuencia de ácido nucleico diana complementaria a la secuencia de unión al ácido nucleico presente en uno de

los polinucleótidos Cas.

Como se utiliza en el presente documento "complementariedad" se refiere a la capacidad de una secuencia de ácido nucleico para formar enlaces de hidrógeno con otra secuencia de ácido nucleico (por ejemplo, mediante el emparejamiento de bases de Watson-Crick tradicional). Un porcentaje de complementariedad indica el porcentaje de restos en una molécula de ácido nucleico que puede formar enlaces de hidrógeno con una segunda secuencia de ácido nucleico. Si dos secuencias de polinucleótido tienen un 100 % de complementariedad, las dos secuencias son perfectamente complementarias, es decir, todos los restos contiguos por enlaces de hidrógeno de un primer polinucleótido con el mismo número de restos contiguos en un segundo polinucleótido.

"Enlace covalente", "fijado covalentemente", "unido covalentemente", "ligado covalentemente", "conectado covalentemente", y "enlace molecular" se utilizan en el presente documento de manera intercambiable, y se refieren a un enlace químico que implica compartir pares de electrones entre los átomos. Ejemplos de enlaces covalentes incluyen, pero no se limitan a, enlaces fosfodiéster y enlaces fosforotioato.

"Enlace no covalente", "fijado no covalentemente", "unido no covalentemente", "ligado no covalentemente", "interacción no covalente", y "conectado no covalentemente" se utilizan en el presente documento de manera intercambiable, y se refieren a cualquier enlace químico relativamente débil que no implica compartir un par de electrones. A menudo múltiples enlaces no covalentes a menudo estabilizan la conformación de macromoléculas y median interacciones entre moléculas. Ejemplos de enlaces no covalentes incluyen, pero no se limitan a enlaces de hidrógeno (por ejemplo, pares de bases unidas por enlaces de hidrógeno tipo Watson-Crick, y emparejamiento de bases Hoogsteen), interacciones iónicas (por ejemplo,  $\text{Na}^+\text{Cl}^-$ ), interacciones de van der Waals, y enlaces hidrófobos.

"Conectar", "conectado" y "conectando" se utilizan de manera intercambiable en el presente documento, y se refiere a un enlace covalente o un enlace no covalente entre dos macromoléculas (por ejemplo, polinucleótidos, proteínas, y similares).

Como se utiliza en el presente documento, "unión" se refiere a una interacción no covalente entre macromoléculas (por ejemplo, entre una proteína y un polinucleótido, entre un polinucleótido y un polinucleótido, y entre una proteína y una proteína, y similares). Dicha interacción no covalente también hace referencia a "asociación" o "interacción" (por ejemplo, si una primera macromolécula interactúa con una segunda macromolécula, la primera macromolécula se une a una segunda macromolécula de una manera no covalente). Algunas partes de una interacción de unión pueden ser específicas de la secuencia (las expresiones "unión específica de secuencia", "unir específicamente la secuencia", y "unión específica del sitio", y "se une específicamente al sitio" se utilizan de manera intercambiable en el presente documento). La unión específica de secuencia, como se utiliza en el presente documento, se refiere normalmente a uno o más NATNA capaces de formar un complejo con una proteína (por ejemplo, una proteína Cas o una proteína Argonauta) para hacer que la proteína se una a una primera secuencia de ácido nucleico (por ejemplo, una primera secuencia de ADN) que comprende una primera secuencia de ácido nucleico diana (por ejemplo, una primera secuencia de ADN diana) preferentemente con respecto a una segunda secuencia de ácido nucleico (por ejemplo, una segunda secuencia de ADN sin la secuencia de unión del ácido nucleico diana (por ejemplo, la primera secuencia de unión del ADN diana). Todos los componentes de una interacción de unión no tienen que ser específicos de secuencia, tal como los contactos de una proteína con restos de fosfato en un armazón de ADN. Las interacciones de unión se pueden caracterizar por una constante de disociación (Kd). "Afinidad" se refiere a la fuerza de unión. Una afinidad de unión aumentada se correlaciona con una Kd más baja.

Como se utiliza en el presente documento, una proteína Cas (por ejemplo, una proteína Cas9 o una proteína Cpf1) se dice que "se dirige" a un polinucleótido si un complejo proteína/NATNA nucleoproteico se une o escinde un polinucleótido en la secuencia de ácido nucleico diana con el polinucleótido. Se dice que un complejo proteína Argonauta/NATNA nucleoproteico se une o escinde un polinucleótido en la secuencia de ácido nucleico diana con el polinucleótido.

Como se utiliza en el presente documento, "rotura de cadena doble" (DSB) se refiere a ambas cadenas de un segmento de ADN de cadena doble que se ha cortado. En algunos casos, si dicha rotura se produce, se puede decir que una cadena tiene un "extremo pegajoso" en el que los nucleótidos están expuestos y no unidos por hidrógeno a los nucleótidos de la otra cadena. En otros casos, se puede producir un "extremo romo" en el que ambas cadenas mantienen completamente emparejadas en sus bases entre ellas a pesar de la DSB.

"Polinucleótido donante", "oligonucleótido donante", y "armazón donante" se utilizan de manera intercambiable en el presente documento y puede ser un polinucleótido de cadena doble (por ejemplo, ADN), un polinucleótido de cadena sencilla (por ejemplo, oligonucleótidos de ADN), o una combinación de los mismos. Los polinucleótidos donantes comprenden brazos de homología que flanquean la secuencia de inserción (por ejemplo, las DSB del ADN). Los brazos de homología de cada lado pueden variar de longitud. Los parámetros para el diseño y construcción de polinucleótidos donantes se conocen bien en la técnica (véase, por ejemplo, Ran, F., et al, Nature Protocols 8(11):2281-2308 (2013); Smithies, O., et al, Nature 317:230-234 (1985); Thomas, K., et al, Cell 44:419-428 (1986); Wu, S., et al, Nature Protocols 3:1056-1076 (2008); Singer, B., et al, Cell 31:25-33 (1982); Shen, P., et al, Genetics 112:441-457 (1986); Watt, V., et al, Proceedings of the National Academy of Sciences of the United States of America 82:4768-4772 (1985),

Sugawara, N., et al, *Journal of Molecular Cell Biology* 12(2):563-575 (1992); Rubnitz, J., et al, *Journal of Molecular Cell Biology* 4(11):2253-2258 (1984); Ayares, D., et al, *Proceedings of the National Academy of Sciences of the United States of America* 83(14):5199-5203 (1986); Liskay, R, et al, *Genetics* 115(1): 161-167 (1987).

5 Como se utiliza en el presente documento, "preparación dirigida por homología" (HDR) se refiere a la reparación de ADN que tiene lugar en las células, por ejemplo, durante la reparación de una DSB en el ADN. La HDR necesita la secuencia de homología del nucleótido y utiliza un polinucleótido donante para reparar la secuencia en el que se produce la DSB (por ejemplo, en una secuencia de ADN diana). El polinucleótido donante en general tiene la secuencia de homología requerida que flanquea la secuencia que flanquea la DSB de manera que el polinucleótido donante puede servir como un almacén adecuado para la reparación. La HDR da como resultado la transferencia de información genética, por ejemplo, desde el polinucleótido donante a la secuencia de ADN diana. La HDR puede resultar en la alteración de la secuencia de ADN diana (por ejemplo, inserción, eliminación, mutación) si el polinucleótido donante se diferencia de la secuencia de ADN diana y se incorpora parte o todo el polinucleótido donante en la secuencia de ADN diana. En algunas realizaciones, todo el polinucleótido donante, una parte del polinucleótido donante, o una copia del polinucleótido donante se integra en el sitio de la secuencia de ADN diana. Por ejemplo, se puede utilizar un polinucleótido donante para la reparación de la rotura de la secuencia de ADN diana, en el que la reparación da como resultado la transferencia de información genética (es decir, secuencias de polinucleótido) desde el polinucleótido donante al sitio o en estrecha proximidad de la rotura del ADN. En consecuencia, la nueva información genética (es decir, las secuencias de polinucleótido) se pueden insertar o copiar en una secuencia de ADN diana.

10 Una "región genómica" es un segmento en un cromosoma del genoma de una célula huésped que está presente en cualquier lado del sitio de la secuencia del ácido nucleico diana o, de manera alternativa, también incluye una parte del sitio de la secuencia del ácido nucleico diana. Los brazos de homología del polinucleótido donante tienen homología suficiente para someterse a recombinación homóloga con las regiones genómicas correspondientes. En algunas realizaciones, los brazos de homología del polinucleótido donante comparten una homología de secuencia significativa con la región genómica que flanquea inmediatamente el sitio la secuencia de ácido nucleico diana; se reconoce que los brazos de homología se pueden diseñar para que tengan suficiente homología con las regiones genómicas más allá del sitio de la secuencia de ácido nucleico diana.

15 20 Como se utiliza en el presente documento, "unión de extremos no homólogos" (NHEJ) se refiere a la reparación de una DSB en el ADN por ligadura directa de un extremo de la rotura al otro extremo de la rotura sin necesidad de un polinucleótido donante. La NHEJ es una ruta de reparación de ADN disponible para que las células reparen el ADN sin el uso de un almacén de reparación. La NHEJ en ausencia de un polinucleótido donante da como resultado a menudo nucleótidos que se insertan o se eliminan aleatoriamente en el sitio de la DSB.

25 30 "Unión de extremos mediada por microhomología" (MMEJ) es una ruta para la reparación de la DSB en el ADN. La MMEJ implica eliminaciones que flanquean una DSB y el alineamiento de secuencias microhomólogas internas a los extremos rotos antes de la unión. La MMEJ se define genéticamente y necesita la actividad, por ejemplo, de CtlP, Poli (ADP-Ribosa) Polimerasa 1 (PARP1), ADN polimerasa theta (Pol  $\theta$ ), ADN Ligasa 1 (Lig 1), DNA Ligasa 3 (Lig 3). Se conocen componentes genéticos adicionales en la técnica (véase, por ejemplo, Sfeir, A., et al, *Trends in Biochemical Sciences* 40:701-714 (2015)).

35 40 Como se utiliza en el presente documento, "reparación de ADN" engloba cualquier proceso en el que la maquinaria celular repara el daño de una molécula de ADN contenida en la célula. El daño reparado puede incluir roturas de cadena sencilla o roturas de cadena doble. Existen al menos tres mecanismos para reparar las DSB: HDR, NHEJ y MMEJ. "Reparación de ADN" también se utiliza en el presente documento para referirse a la reparación de ADN que resulta de la manipulación humana, en la que el locus diana se modifica, por ejemplo, por inserción, eliminación, sustitución de nucleótidos, que representan formas de edición genómica.

45 50 Como se utiliza en el presente documento, "recombinación" se refiere a un procedimiento de intercambio de información genética entre dos polinucleótidos.

55 Como se utiliza en el presente documento, las expresiones "secuencias reguladoras", "elementos reguladores", y "elementos de control" son intercambiables y se refieren a secuencias de polinucleótidos que están corriente arriba (secuencias no codificantes 5'), en, o corriente abajo (secuencias no traducidas 3') de un polinucleótido diana que se va a expresar. Las secuencias reguladoras tienen influencia, por ejemplo, en el tiempo de transcripción, cantidad o nivel de transcripción, procesamiento o estabilidad del ARN, y/o traducción de la secuencia de nucleótidos estructural relacionada. Las secuencias reguladoras pueden incluir secuencias de unión al activador, amplificadores, intrones, secuencias de reconocimiento de poliadenilación, promotores, sitios de inicio de la transcripción, secuencias de unión al represor, estructuras tallo-lazo, secuencias de inicio de la traducción, sitios de entrada al ribosoma interno (IRES), secuencias líder de traducción, secuencias de terminación de la traducción, sitios de unión a cebadores, y similares.

60 65 Los elementos reguladores incluyen los que dirigen la expresión constitutiva, inducible, y represible de una secuencia de nucleótidos en muchos tipos de tipos de células huésped y los que dirigen la expresión de la secuencia de nucleótidos solo en ciertas células huésped (por ejemplo, secuencias reguladoras específicas de tejido). En algunas realizaciones, un vector comprende uno o más promotores pol III, uno o más promotores pol II, uno o más promotores

pol I, o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, por no se limitan a los promotores U6 y HI. Ejemplos de promotores pol II incluyen, pero no se limitan, a un promotor LTR del virus retroviral de sarcoma de Rous (RSV) (opcionalmente con el amplificador del RSV), el promotor del citomegalovirus (CMV) (opcionalmente con el amplificador de CMV; véase, por ejemplo, Boshart et al, Cell 41:521-530 (1985)), el promotor de SV40, el promotor dihidrofolato reductasa, el promotor P-actina, el promotor fosfoglicerol cinasa (PGK), y el promotor EFla. Los expertos en la técnica apreciarán que el diseño de un vector de expresión puede depender de factores tales como la elección de la célula huésped que se va a transformar, el nivel de expresión deseado, y similares. Se puede introducir un vector en células huésped para producir de esta manera transcripciones, proteínas, o péptidos, incluyendo proteínas o péptidos de fusión, codificados por ácidos nucleicos como se describe en el presente documento.

“Gen” como se utiliza en el presente documento, se refiere a una secuencia de polinucleótido que comprende exones y secuencias reguladoras relacionadas. Un gen puede comprender adicionalmente intrones y/o regiones no traducidas (UTR).

Como se utiliza en el presente documento, la expresión “unida operativamente” se refiere a secuencias de polinucleótido o secuencias de aminoácidos colocados en una relación funcional entre ellas. Por ejemplo, las secuencias reguladoras (por ejemplo, un promotor o amplificador) están “unidas operativamente” a un polinucleótido que codifica un producto genético si las secuencias reguladoras regulan o contribuyen a la modulación de la transcripción del polinucleótido. Los elementos reguladores unidos operativamente normalmente están contiguos a la secuencia codificante, sin embargo, los amplificadores pueden funcionar si se separan de un promotor hasta varias kilobases o más. En consecuencia, algunos elementos reguladores pueden unirse operativamente a una secuencia de polinucleótido, pero no están contiguos a la secuencia de polinucleótido. De manera similar, los elementos reguladores traduccionales contribuyen a la modulación de la expresión proteica a partir de un polinucleótido.

Como se utiliza en el presente documento, “expresión” se refiere a la transcripción de un polinucleótido a partir de un armazón de ADN, que da como resultado, por ejemplo, un ARN mensajero (ARNm) u otra transcripción de ARN (por ejemplo, no codificante, tal como ARN estructurales o de armazón). El término adicional se refiere al proceso mediante el que el ARN m transcrito se traduce en péptidos, polipéptidos o proteínas. Se puede hacer referencia a las transcripciones y polipéptidos codificados colectivamente como “productos genéticos”. La expresión puede incluir el corte y empalme del ARNm en una célula eucariota, si el polinucleótido se deriva de ADN genómico.

Como se utiliza en el presente documento, el término “modular” se refiere a un cambio en la cantidad, grado o cantidad de una función. Por ejemplo, un complejo proteína Cas9/proteína Cpf1/Cas9-NATNA-Cpf1, como se desvela en el presente documento, se puede modular la actividad de una secuencia promotora uniendo las dos secuencias de ácido nucleico diana en o cerca del promotor. Dependiendo de la acción que se produce después de la unión, el complejo proteína Cas9/proteína Cpf1/Cas9-NATNA-Cpf1 puede inducir, aumentar, suprimir, o inhibir la transcripción de un gen unido operativamente a la secuencia promotora. Por lo tanto “modulación” de la expresión genética incluye tanto la activación genética como la represión genética.

La modulación se puede ensayar para determinar cualquier característica que esté afectada directa o indirectamente por la expresión del gen diana. Dichas características incluyen, por ejemplo, cambios en los niveles de ARN o proteína, actividad proteica, niveles del producto, expresión del gen, o nivel de actividad de genes indicadores. En consecuencia, las expresiones “modulación de la expresión”, “inhibición de la expresión”, y “activación de la expresión” de un gen se puede referir a la capacidad de un complejo proteína Cas9/proteína Cpf1/Cas9-NATNA-Cpf1 para cambiar, activar o inhibir la transcripción de un gen.

“Vector” y “plásmido”, como se utilizan en el presente documento, se refieren a un vehículo del polinucleótido para introducir material genético en una célula. Los vectores pueden ser lineales o circulares. Los vectores pueden contener una secuencia de replicación capaces de efectuar la replicación del vector en una célula huésped adecuada (es decir, un origen de replicación). Al transformar un huésped adecuado, el vector puede replicarse y funcionar independientemente del genoma del huésped o integrarse en el genoma del huésped. El diseño del vector depende, entre otras cosas, del uso que se pretende y la célula huésped del vector, y el diseño de un vector de la invención para un uso y célula huésped particulares está en la experiencia de la técnica. Los cuatro tipos principales de vectores son plásmidos, vectores víricos, cósmidos, y cromosomas artificiales. Normalmente, los vectores comprenden un origen de replicación, un sitio de multiclonación, y/o un marcador genético. Un vector de expresión normalmente comprende un casete de expresión.

Como se utiliza en el presente documento, “casete de expresión” se refiere a una construcción de polinucleótido que se genera utilizando métodos recombinantes o por medios sintéticos y que comprenden secuencias reguladoras unidas operativamente a un polinucleótido seleccionado para facilitar la expresión del polinucleótido seleccionado en una célula huésped. Por ejemplo, las secuencias reguladoras pueden facilitar la transcripción del polinucleótido seleccionado en una célula huésped, o la transcripción y traducción del polinucleótido seleccionado en una célula huésped. Un casete de expresión puede, por ejemplo, integrarse en el genoma de una célula huésped o puede estar presente en un vector para formar un vector de expresión.

Como se utiliza en el presente documento, un “vector de direccionamiento” es una construcción de ADN recombinante

- que comprende normalmente brazos de ADN a medida, homólogos de ADN genómico, que flanquea elementos de un gen diana o secuencia de ácido nucleico diana (por ejemplo, un DSB). Un vector de direccionamiento comprende un polinucleótido donante. Los elementos del gen diana se pueden modificar de distintas maneras incluyendo eliminaciones y/o inserciones. Un gene diana defectuoso se puede sustituir por un gen diana funcional, o de manera alternativa un gen funcional puede inactivarse. Opcionalmente, el polinucleótido donante de un vector de direccionamiento comprende un casete de selección que comprende un marcador genético que se introduce en el gen diana. Las regiones de direccionamiento (es decir, las secuencias de ácido nucleico diana) adyacentes o en un gen diana se pueden utilizar para afectar la regulación de la expresión del gen.
- 10 Como se utiliza en el presente documento, las expresiones “ácido nucleico”, “secuencia de nucleótidos”, “oligonucleótido”, y “polinucleótido” son intercambiables y se refieren a una forma polimérica de nucleótidos. Los nucleótidos pueden ser desoxirribonucleótidos (ADN) o ribonucleótidos (ARN), análogos de los mismos, o combinaciones de los mismos y pueden tener cualquier longitud. Los polinucleótidos pueden llevar a cabo cualquier función y pueden tener cualquier estructura secundaria y terciaria. Las expresiones engloban análogos conocidos de nucleótidos naturales y nucleótidos que se han modificado en la base, azúcar y restos de fosfato. Los análogos de un nucleótido particular tienen la misma especificidad de emparejamiento de bases (por ejemplo, un análogo de la base A se empareja con T). Un polinucleótido puede comprender un nucleótido modificado o múltiples nucleótidos modificados. Ejemplos de nucleótidos modificados incluyen nucleótidos fluorados, nucleótidos metilados, y análogos de nucleótidos. La estructura de nucleótidos se puede modificar antes o después de ensamblar el polímero. Después de la polimerización, se pueden modificar adicionalmente los polinucleótidos mediante, por ejemplo, la conjugación con un componente marcador o componente de unión a la diana. Una secuencia de nucleótido puede incorporar componentes no nucleotídicos. Las expresiones también engloban ácidos nucleicos que comprenden restos de armazón o enlaces modificados, que son sintéticos, de origen natural, y de origen no natural, y tienen propiedades de unión similares que el polinucleótido de referencia (por ejemplo, ADN o ARN). Ejemplos de dichos análogos incluye, pero no se limitan a, fosforotioatos, fosforoamidatos, metil fosfonatos, metil fosfonatos quirales, 2-O-metil ribonucleótidos, ácidos nucleicos peptídicos (PNA), nucleósidos de ácidos nucleicos bloqueados (LNA™) (Exigon, Inc., Woburn, MA), ácido nucleico glicol, ácidos nucleicos puenteados, y estructuras morfolino.
- 15 Los ácidos nucleicos peptídicos (PNA) son homólogos sintéticos de ácidos nucleicos en los que el armazón polinucleotídico fosfato-azúcar se sustituye por un polímero pseudo-peptídico flexible. Las bases nucleicas están unidas al polímero. Los PNA tienen la capacidad para hibridarse con alta afinidad y especificidad a secuencias complementarias de ARN y ADN.
- 20 En los ácidos nucleicos con fosforotioato, el enlace de fosforotioato (PS) sustituye un átomo de azufre por un oxígeno sin puenteo en la matriz de polinucleótido con fosfato. Esta modificación hace que el enlace internucleotídico sea resistente a la degradación por nucleasas. En algunas realizaciones los enlaces fosforotioato se introducen entre el último nucleótido 3 a 5 en el extremo 5' o 3' de una secuencia de polinucleótido para inhibir la degradación por exonucleasas. La colocación de enlaces fosforotioato a lo largo del oligonucleótido completo ayuda también a reducir la degradación por endonucleasas.
- 25 Los ácidos nucleicos treosa (ATN) son polímeros genéticos artificiales. La estructura del armazón del ATN comprende una repetición de azúcares treosa unidas por enlaces fosfodiéster. Los polímeros ARN son resistentes a la degradación por nucleasas. El ATN puede auto-ensamblarse por enlaces de hidrógeno de pares de bases en estructuras dúplex.
- 30 Se pueden introducir inversiones de enlace en los polinucleótidos mediante el uso de “fosforoamiditas invertidas” (véase, por ejemplo, [www.ucalgary.ca/dnalab/synthesis/-modifications/linkages](http://www.ucalgary.ca/dnalab/synthesis/-modifications/linkages)). Normalmente, dichos polinucleótidos tiene grupos fosforoamidita en la posición 5'-OH y un grupo protector dimetoxitritil (DMT) en la posición 3'-OH. Normalmente, el grupo protector DMT está en el 5'-OH y la fosforoamidita está en el 3'-OH. El uso más común de la inversión de enlaces es añadir un enlace 3'-3' al final de un polinucleótido con una estructura con fosforotioato. El enlace 3'-3' estabiliza el polinucleótido frente a la degradación por exonucleasas creando un oligonucleótido que tiene dos extremos 5'-OH y sin extremo 3'-OH.

Las secuencias de polinucleótidos se exponen en el presente documento en orientación convencional 5' a 3' a menos de que se indique otra cosa.

5 Como se utiliza en el presente documento, "identidad de secuencia" se refiere en general al porcentaje de identidad de bases de nucleótidos o aminoácidos en comparación con un primer polinucleótido o polipéptido con respecto a un segundo polinucleótido o polipéptido utilizando algoritmos que tienen distintos parámetros de valoración. La identidad de secuencia entre dos polinucleótidos o dos polipéptidos se puede determinar utilizando el alineamiento de secuencia por distintos métodos y programas de computadora (por ejemplo, BLAST, CS-BLAST, FASTA, HMMER, L-ALIGN, y similares) disponibles mediante internet en los sitios que incluyen pero no se limitan a GENBANK 10 ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) y EMBL-EBI ([www.ebi.ac.uk](http://www.ebi.ac.uk)). La identidad de secuencia entre dos polinucleótidos o dos polipéptidos se calcula generalmente utilizando los parámetros por defecto de referencia de distintos métodos o programas de computadora. Un alto grado de identidad de secuencia, como se utiliza en el presente documento, entre dos polinucleótidos o dos polipéptidos está normalmente entre aproximadamente un 90 % de identidad y un 100 % de identidad, por ejemplo, aproximadamente un 90 % de identidad o más, preferentemente aproximadamente 95 % o 15 más de identidad, más preferentemente aproximadamente un 98 % o más de identidad. Un grado moderado de identidad de secuencia, como se utiliza en el presente documento, entre dos polinucleótidos o dos polipéptidos está normalmente entre aproximadamente un 80 % > de identidad a aproximadamente un 85 % de identidad, por ejemplo, aproximadamente un 80 % o más de identidad, preferentemente aproximadamente un 85 % de identidad. Un grado bajo de identidad de secuencia, como se utiliza en el presente documento, entre dos polinucleótidos o dos polipéptidos está normalmente entre aproximadamente un 50 % de identidad y un 75 % de identidad, por ejemplo, aproximadamente un 50 % de identidad, preferentemente aproximadamente un 60 % de identidad, más preferentemente aproximadamente un 75 % de identidad. Por ejemplo, una proteína Cas (por ejemplo, una Cas9 comprende sustituciones de aminoácidos o una CpfI comprende sustituciones de aminoácidos) puede tener un grado moderado de identidad de secuencia, o preferentemente un alto grado de identidad de secuencia, a lo largo de su longitud con respecto a una proteína Cas de referencia (por ejemplo, una Cas9 de tipo silvestre o una CpfI de tipo silvestre, respectivamente). Como otro ejemplo, un NATNA puede tener un grado moderado de identidad de secuencia, o preferentemente un alto grado de identidad de secuencia, a lo largo de su longitud en comparación con un polinucleótido de tipo silvestre de referencia que forma un complejo con la proteína Cas de referencia (por ejemplo, un sgARN que forma un complejo con Cas9 o un crARN que forma un complejo con CpfI).

30 Como se utiliza en el presente documento, "hibridación" o "hibridar" o "que se hibrida" es el proceso de combinación de dos moléculas de ADN de cadena doble o ARN complementarias de manera que se forma una única molécula de cadena doble (ADN/ADN, ADN/ARN, ARN/ARN) mediante emparejamiento de bases por hidrógeno. La rigurosidad de hibridación se determina normalmente por la temperatura de hibridación y la concentración de sales del tampón de hibridación; por ejemplo, a alta temperatura y baja sal se proporcionan condiciones de hibridación con alta rigurosidad. Ejemplos de intervalos de concentración de sales e intervalos de temperatura para diferentes condiciones de hibridación son los siguientes: alta rigurosidad, aproximadamente 0,01 M a aproximadamente 0,05 M de sales, temperatura de hibridación de 5 °C y 10 °C por debajo de la  $T_m$ ; la rigurosidad moderada, aproximadamente 0,15 M a aproximadamente 0,33 M de sal, temperatura de hibridación de 20 °C a 29 °C por debajo de la  $T_m$ ; y baja rigurosidad, aproximadamente 0,33 M a aproximadamente 0,82 M de sal, temperatura de hibridación de 40 °C a 48 °C por debajo de la  $T_m$ . La  $T_m$  del dúplex de ácidos nucleicos se calcula por métodos convencionales bien conocidos en la técnica (véase, por ejemplo, Maniatis, T., et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press: New York (1982); Casey, J., et al, Nucleic Acids Research 4:1539-1552 (1977); Bodkin, D.K., et al, Journal of Virological Methods 10(1):45-52 (1985); Wallace, R.B., et al, Nucleic Acids Research 9(4):879-894 (1981)). Las herramientas del algoritmo de predicción para estimar la  $T_m$  también están ampliamente disponibles. Las condiciones de alta rigurosidad para la hibridación normalmente se refieren a condiciones en las cuales un ácido nucleico que tiene complementariedad con una secuencia diana se hibrida predominantemente con la secuencia diana, y sustancialmente no se hibrida con secuencias no dianas. Normalmente, las condiciones de hibridación son de rigurosidad moderada, preferentemente rigurosidad alta.

50 Como se utiliza en el presente documento, una "estructura tallo-lazo" o un "elemento tallo-lazo" se refiere a un polinucleótido que tiene una estructura secundaria que incluye una región de nucleótidos que se conocen o se ha previsto que forme una región de cadena doble (el "elemento tallo"), en el que en un extremo de la región de cadena doble de la región de cadena doble está unida por una región de nucleótidos predominantemente de cadena sencilla (el "elemento lazo"). El término elemento "ahorquillado" también se utiliza en el presente documento para referirse a las estructuras tallo-lazo. Dichas estructuras son bien conocidas en la técnica. El emparejamiento de bases puede ser exacto; sin embargo, como se conoce en la técnica, un elemento tallo no necesita un emparejamiento de bases exacto. Por lo tanto, el elemento tallo puede incluir una o más faltas de coincidencia o bases no emparejadas.

60 Una "secuencia de nucleótidos de elemento enlazador" y una "secuencia de nucleótidos enlazadora" se utilizan de manera intercambiable en el presente documento y normalmente se refiere a una secuencia de uno o más nucleótidos unidos covalentemente a un extremo 5', un extremo 3', o ambos extremos 5' y 3' de una primera secuencia de polinucleótido. En algunas realizaciones, la secuencia de nucleótidos de elemento enlazador tiene el fin de conectar un primer polinucleótido a un segundo polinucleótido.

65 Como se utiliza en el presente documento, el término "aminoácido" se refiere a aminoácidos naturales y sintéticos (no

naturales), incluyendo análogos de aminoácidos, aminoácidos modificados, peptidomiméticos, glicina, e isómeros D o L.

5 Como se utiliza en el presente documento, los términos “péptido”, “polipéptido”, y “proteína” son intercambiables y se refieren a polímeros de aminoácidos. Un polipéptido puede tener cualquier longitud. Puede ser ramificado o lineal, puede estar interrumpido por no aminoácidos, y puede comprender aminoácidos modificados. Los términos se pueden utilizar para hacer referencia a un polímero de aminoácidos que se ha modificado mediante, por ejemplo, acetilación, formación de enlaces disulfuro, glicosilación, lipidación, fosforilación, pegilación, biotilación, entrecruzamiento, y/o conjugación (por ejemplo, con un componente marcador o ligando). Las secuencias de polipéptidos se exponen en el presente documento en la orientación del extremo N o extremo C convencional.

15 Los polipéptidos y polinucleótidos se pueden fabricar utilizando técnicas de rutina en el campo de la biología molecular (véase, por ejemplo, los textos convencionales expuestas anteriormente). Además, esencialmente cualquier polipéptido o polinucleótido está disponible en fuentes comerciales.

20 Las expresiones “proteína de fusión” y “proteína quimérica”, como se utiliza en el presente documento, se refiere a una única proteína creada por la unión de dos o más proteínas, dominios proteicos, o fragmentos de proteínas que no existen juntos naturalmente en una única proteína. Por ejemplo, una proteína de fusión puede contener un primer dominio de una proteína Cas9 y un segundo dominio de Cpfl, un primer dominio de una proteína Cas9 o Cpfl y un segundo dominio de una proteína Argonauta, un primer dominio de una proteína Cas9, Cpfl o Argonauta y un segundo dominio de una proteína distinta de una proteína Cas9, Cpfl o Argonauta, y así. La modificación para incluir dichos dominios en la proteína de fusión puede conferir una actividad adicional en los polipéptidos dirigidos al sitio modificados. Dichas actividades pueden incluir actividad nucleasa, actividad metiltransferasa, actividad desmetilasa, actividad de reparación del ADN, actividad de daño al ADN, actividad de desaminación, actividad dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad de formación de pirimidinas diméricas, actividad integrasa, actividad transposasa, actividad recombinasa, actividad polimerasa, actividad ligasa, actividad helicasa, actividad fotoliasa, actividad glicosilasa, actividad acetiltransferasa, actividad desacetilasa, actividad cinasa, actividad fosfatasa, actividad ubiquitina ligasa, actividad desubiquitinante, actividad de adenilación, actividad de desadenilación, actividad SUMOilante, actividad desSUMOilante, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación o actividad de desmiristoilación que modifica un polipéptido asociado con una secuencia de ácido nucleico diana (por ejemplo, una histona). Una proteína de fusión también puede comprender marcadores epítopos (por ejemplo, marcadores de histidina, marcadores FLAG® (Sigma Aldrich, St. Louis, MO), marcadores Myc), secuencias proteicas indicadoras (por ejemplo, glutatión-S-transferasa, beta-galactosidasa, luciferasa, proteína fluorescente verde, proteína fluorescente azul, proteína fluorescente amarilla), y/o dominios de unión de ácido nucleico (por ejemplo, dominio de unión de ADN, dominio de unión de ARN). Una proteína de fusión puede comprender también dominios activadores (por ejemplo, factores de transcripción de choque térmico, activadores NFκB) o dominios represores (por ejemplo, un dominio KRAB). Como describen Lupo, A., et al, Current Genomics 14(4): 268-278 (2013), el dominio KRAB es un potente módulo de represión transcripcional y se localiza en la secuencia del extremo amino de la mayoría de las proteínas C2H2 en dedos de zinc (véase, por ejemplo, Margolin, J., et al, Proceedings of the National Academy of Sciences of the United States of America 91:4509-4513 (1994); Witzgall, R., et al., Proceedings of the National Academy of Sciences of the United States of America 91:4514-4518 (1994)). El dominio KRAB normalmente se une a proteínas co-represoras y/o factores de transcripción mediante interacciones proteína-proteína, que producen una represión de la transcripción de genes a los que se unen las proteínas en dedos de zinc KRAB (KRAB-ZFP) (véase, por ejemplo, Friedman JR, et al, Genes & Development 10:2067-2678 (1996)). En algunas realizaciones, las secuencias de ácido nucleico enlazadoras se utilizan para unir las dos o más proteínas, dominios proteicos, o fragmentos de proteínas.

50 Como se utiliza en el presente documento, un “célula huésped” se refiere en general a una célula biológica. Una célula es la unidad estructural básica, funcional y/o biológica de un organismo vivo. Una célula puede originarse de cualquier organismo que tenga una o más células. Ejemplos de células huésped incluyen, pero no se limitan a: una célula procariota, célula eucariota, célula bacteriana, célula arquea, una célula de un organismo eucariota unicelular, una célula de protozoo, una célula de una planta (por ejemplo, células de cultivos vegetales (tales como soja, tomates, remolacha azucarera, calabaza, heno, cannabis, tabaco, plátanos, boniato, batata, yuca, patatas, trigo, sorgo, soja, arroz, maíz, choclo, Brassica productora de aceite (por ejemplo, colza productora de aceite y canola), algodón, caña de azúcar, girasol, mijo y alfalfa), frutas, verduras, granos, semillas, plantas de flor, coníferas, gimnospermas, helechos, musgos, líquenes), y células de alga (por ejemplo, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorellapyrenoidosa*, *Sargassumpatens*, *C. agardh*, y similares), algas marinas (por ejemplo, kelp), una célula fúngica (por ejemplo, una célula de levadura, una célula de un hongo), una célula animal, una célula de un animal invertebrado (por ejemplo, una mosca de la fruta, nidario, equinodermo, nematodo, y similares), una célula de un animal vertebrado (por ejemplo, un pez, anfibio, reptil, ave, mamífero), una célula de un mamífero (por ejemplo, un cerdo, una vaca, una cabra, una oveja, un roedor, una rata, un ratón, un primate no humano, un ser humano, y similares). Además, una célula puede ser una célula madre o una célula progenitora.

Como se utiliza en el presente documento, una “célula madre” se refiere a una célula que tiene la capacidad de autorrenovarse, es decir, la capacidad de pasar por numerosos ciclos de división celular mientras que mantiene el estado no diferenciado. Las células madre pueden ser pleni-potenciales, pluripotenciales, multipotenciales, oligopotenciales, o unipotenciales. Las células madre pueden ser embrionarias, fetales, amnióticas, de adulto, o células madre pluripotenciales inducidas.

Como se utiliza en el presente documento, “células madre pluripotenciales inducidas” se refiere a un tipo de célula madre pluripotencial que se deriva artificialmente de una célula no pluripotencial, normalmente una célula somática de adulto, induciendo la expresión de genes específicos.

“Planta”, como se utiliza en el presente documento, se refiere a plantas completas, órganos vegetales, tejidos vegetales, plasma germinal, semillas, células vegetales, y progenie de las mismas. Las células vegetales incluyen, sin limitación, las células de semillas, los cultivos en suspensión, embriones, regiones meristemáticas, tejidos del callo, hojas, raíces, brotes, gametofitos, esporofitos, polen y microsporas. Las partes de plantas incluyen tejidos diferenciados y no diferenciados que incluyen, por no se limitan a raíces, tallos, brotes, hojas, pólenes, semillas, tejidos tumorales y distintas formas de células y cultivos (por ejemplo, células únicas, protoplastos, embriones y tejido del callo). el tejido vegetal puede estar en la planta o en un órgano, tejido o cultivo celular vegetal. “Órgano vegetal” se refiere a un tejido vegetal o grupo de tejidos que constituyen una parte distintiva morfológica y funcionalmente de una planta.

“Sujeto”, como se utiliza en el presente documento, se refiere a cualquier miembro del Filum Cordata, que incluye, sin limitación, seres humanos y otros primates, incluyendo primates no humanos tales como macacos rhesus, chimpancés y otras especies de monos y simios; animales de granja, tales como el ganado bovino, ovejas, cerdos, cabras, y caballos; mamíferos domésticos, tales como perros y gatos; animales de laboratorio que incluyen conejos, ratones, ratas y cobayas; aves, que incluyen aves domésticas, silvestres y de caza, tales como pollos, pavos y otras aves gallináceas, patos y gansos; y similares. El término no denota una edad o género en particular. Por lo tanto, se pretende que estén cubiertos individuos adultos, jóvenes, y recién nacidos, así como machos y hembras. En algunas realizaciones, una célula huésped se deriva de un sujeto (por ejemplo, células madre, células progenitores, células específicas de tejido). En algunas realizaciones, el sujeto es un sujeto no humano.

Las expresiones “de tipo silvestre”, “de origen natural”, y “no modificado” como se utiliza en el presente documento significa la forma, apariencia, fenotipo o cepa típica (o la más común) que existe en la naturaleza; por ejemplo, la forma típica de las células, organismos, características, polinucleótidos, proteínas, complejos macromoleculares, genes, ARN, ADN, o genomas según existen, y se pueden aislar de una fuente en la naturaleza. La forma apariencia, fenotipo o cepa de tipo silvestre del parental original antes de una modificación intencionada. Por lo tanto, las formas mutantes, variantes, modificadas, recombinantes y alteradas no son formas de tipo silvestre.

Como se utiliza en el presente documento, las expresiones “modificado”, “modificado genéticamente”, “recombinante”, “alterado” y “no de origen natural” son intercambiables e indican una modificación humana intencionada.

Como se utiliza en el presente documento, un “organismo transgénico” se refiere a un organismo cuyo genoma se ha modificado genéticamente. La expresión incluye la progenie (cualquier generación) de un organismo transgénico, a condición de que la progenie tenga la modificación genética.

Como se utiliza en el presente documento, “aislado” puede referirse a un ácido nucleico o polipéptido que, mediante la intervención humana, existe fuera de su ambiente nativo y por lo tanto no es un producto de la naturaleza. Un ácido nucleico o polipéptido aislado puede existir en forma purificada y/o en un ambiente no nativo tal como, por ejemplo, en una célula recombinante.

En un aspecto se desvela un ácido nucleico de tipo cruzado modificado que se dirige a un ácido nucleico (“NATNA de tipo cruzado”). El NATNA de tipo cruzado comprende normalmente un primer NATNA que se dirige a una primera proteína para unirse específicamente del sitio a un a primera secuencia de ácido nucleico diana conectado con un segundo NATNA que se dirige a una segunda proteína para unirse específicamente del sitio a una segunda secuencia de ácido nucleico diana. La conexión puede ser covalente o no covalente La primera proteína comprende preferentemente uno o más dominios nucleasa catalíticamente activos y/o uno o más dominios de nucleasa catalíticamente inactivos. La segunda proteína comprende preferentemente uno o más dominios de nucleasa catalíticamente activos y/o uno o más dominios de nucleasa catalíticamente inactivos. En algunos aspectos el NATNA de tipo cruzado comprende un primar NATNA CRISPR Cas de un primer tipo de sistema CRISPR Cas (por ejemplo, un Cas-NATNA es capaz de formar un complejo con una proteína Cas de un sistema CRISPR de tipo I, II, III, IV, V, o VI) conectado con un segundo tipo de NATNA CRISPR Cas de un segundo sistema CRISPR Cas (por ejemplo, un Cas-NATNA que es capaz de formar un complejo con una proteína Cas de un sistema CRISPR Tipo, I, II, III, IV, V o VI diferente del primer sistema CRISPR Cas.

En otros aspectos, el NATNA tipo cruzado comprende un primer NATNA CRISPR-Cas de un sistema CRISPR Cas Clase 1 (por ejemplo, un Cas-NATNA que es formar un complejo con una proteína Cas de un sistema CRISPR de Tipo I, III, o IV) conectado con un segundo NATNA CRISPR Cas de un sistema CRISPR de Clase 2 (por ejemplo, un

Cas-NATNA que es capaz de formar un complejo con una proteína Cas de un sistema CRISPR de Tipo II, V, o VI). En aspectos adicionales, el NATNA de tipo cruzado comprende un primer NATNA CRISPR de un sistema CRISPR Cas de Clase 1 (por ejemplo, un Cas-NATNA que es capaz de formar un complejo con una proteína Cas de un sistema CRISPR de tipo I, III, o IV) conectado con un segundo NATNA CRISPR Cas diferente de un sistema CRISPR Cas de Clase 1 (por ejemplo, un Cas-NATNA que es capaz de formar un complejo con una proteína Cas de un sistema CRISPR de Tipo I, III, o IV). En aspectos adicionales, un NATNA de tipo cruzado comprende un Ago NATNA y cualquiera de un NATNA con CRISPR de Clase 1 o un NATNA con CRISPR de Clase 2.

Una realización preferida de la presente invención incluye un ácido nucleico de tipo cruzado con CRISPR de clase 2 modificado que se dirige a un ácido nucleico ("NATNA tipo cruzado con CRISPR de Clase 2"), que comprende un Cpfl-NATNA, un primer Cas9-NATNA, conectado con un segundo Cas9-NATNA en el que el primer Cas9-NATNA o el segundo Cas9-NATNA se conecta covalente o no covalentemente con el Cpfl-NATNA.

En algunas realizaciones, se conectan covalentemente dos NATNA diferentes para formar un NATNA de tipo cruzado. Por ejemplo, un primer Cas9-NATNA o un segundo Cas9-NATNA se conectan covalentemente con un Cpfl-NATNA y el Cpfl-NATNA se une covalentemente al extremo 5' o el extremo 3' del primer Cas9-NATNA o el extremo 5' o el extremo 3' del segundo Cas9-NATNA. En realizaciones adicionales, los NATNA diferentes se unen covalentemente mediante una secuencia de nucleótidos de elemento enlazador. Por ejemplo, un primer Cas9-NATNA o un segundo Cas9-NATNA se conecta covalentemente con un Cpfl-NATNA mediante una secuencia de nucleótidos de elemento enlazador, y el Cpfl-NATNA está unido covalentemente mediante una secuencia de nucleótidos de elemento enlazador al extremo 5' o el extremo 3' del primer Cas9-NATNA o el extremo 5' o el extremo 3' del segundo Cas9-NATNA.

Las FIG. 8, FIG. 9, y FIG. 10 ilustran cada una un ejemplo de dichas uniones covalentes utilizando los ARN NATNA.

La FIG. 8 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 5' de un Cpfl-crARN está unido covalentemente a un extremo 3' de un Cas9-crARN. En la figura, la flecha indica la región de unión entre el Cpfl-crARN y el Cas9-crARN. El Cpfl-crARN y el Cas9-crARN pueden unirse directamente entre ellos por un enlace covalente o, por ejemplo, se unen entre ellos mediante una secuencia de elemento enlazador (por ejemplo, el extremo 5' de un Cpfl-crARN se une covalentemente al extremo 3' de una secuencia de nucleótidos de elemento enlazador y el extremo 5' de la secuencia de nucleótidos de elemento enlazador se unen al extremo 3' de un Cas9-crARN).

La FIG. 9 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 3' de un Cpfl-crARN se une covalentemente al extremo 5' de un Cas9-tracrARN. En la figura, la flecha indica la región de unión entre el Cpfl-crARN y el Cas9-tracrARN. El Cpfl-crARN y el Cas9-tracrARN se pueden unir directamente entre ellos por un enlace covalente o, por ejemplo, unirse entre ellos mediante una secuencia de nucleótidos de elemento enlazador (por ejemplo, el extremo 3' de un Cpfl-crARN se une covalentemente al extremo 5' de una secuencia de nucleótidos de elemento enlazador y el extremo 3' de la secuencia de nucleótidos del elemento enlazador se une covalentemente al extremo 5' de un Cas9-tracrARN).

La FIG. 10 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 3' de un Cpfl-crARN se une covalentemente al extremo 5' de un Cas9-crARN. En la figura, la flecha indica la región de unión entre el Cpfl-crARN y el Cas9-crARN. El Cpfl-crARN y el Cas9-crARN se pueden unir directamente entre ellos por un enlace covalente o, por ejemplo, unirse entre ellos mediante una secuencia de nucleótidos de elemento enlazador (por ejemplo, el extremo 3' de un Cpfl-crARN se une covalentemente al extremo 5' de una secuencia de nucleótidos de elemento enlazador y el extremo 3' de la secuencia de nucleótidos del elemento enlazador se une covalentemente al extremo 5' de un Cas9-crARN).

En algunas realizaciones, el NATNA de tipo cruzado comprende dos NATNA diferentes, en el que uno de los NATNA comprende más de un polinucleótido (por ejemplo, FIG. 8, FIG. 9, FIG. 10). El NATNA comprende más de un polinucleótido que se puede unir covalentemente mediante un elemento bucle para formar un NATNA sencillo (por ejemplo, compárese la FIG. 1B, un Cas9-crARN/Cas9-tracrARN con la FIG. 2, un Cas9-sgARN).

Por ejemplo, el extremo 3' de un primer Cas9-NATNA se puede conectar covalentemente mediante un elemento bucle con el extremo 5' de un segundo Cas9-NATNA formando un ácido nucleico asociado con Cas9 que se dirige a un ácido nucleico sencillo ("Cas9-NATNA sencillo") que tiene un extremo 5' y un extremo 3'. Por lo tanto, el Cas9-NATNA sencillo comprende el primer Cas9-NATNA y el segundo Cas9-NATNA. El Cas9-NATNA sencillo está unido covalentemente a un Cpfl-NATNA para formar un NATNA de tipo cruzado con CRISPR de Clase 2. La FIG. 4, FIG. 6 y FIG. 7 ilustran cada una un ejemplo de dichas uniones covalentes utilizando ARN NATNA.

La FIG. 4 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 5' de un Cpfl-crARN está unido covalentemente al extremo 3' de un Cas9-sgARN. En la figura, la flecha indica la región de unión entre el Cpfl-crARN y el Cas9-sgARN. El Cpfl-crARN y el Cas9-sgARN se pueden unir directamente entre ellos por un enlace covalente o, por ejemplo, unirse entre ellos mediante una secuencia de nucleótidos de elemento enlazador (por ejemplo, el extremo 5' de un Cpfl-crARN se une covalentemente al extremo 3' de una secuencia de nucleótidos de elemento enlazador y el extremo 5' de la secuencia de nucleótidos del elemento enlazador está unido

covalentemente al extremo 3' de un Cas9-sgARN).

La FIG. 6 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 5' de un Cpfl-crARN está unido covalentemente al extremo 3' de un Cas9-sgARN, y en el que uno de los elementos ahorquillados 3' (por ejemplo, FIG. 2, 207, 208) se ha eliminado (es decir, borrado) del Cas9-sgARN. En la figura, la flecha indica la región de unión entre el Cpfl-crARN y el Cas-sgARN. El Cpfl-crARN y el Cas9-sgARN pueden unirse directamente entre ellos por un enlace covalente o, por ejemplo, unirse entre ellos mediante una secuencia de nucleótidos de elemento enlazador (por ejemplo, el extremo 5' de un Cpfl-crARN se une covalentemente al extremo 3' de una secuencia de nucleótido de elemento enlazador y el extremo 5' de la secuencia de nucleótidos del elemento enlazador se une covalentemente al extremo 3' de un Cas9-sgARN, en el que uno de las horquillas 3' se ha eliminado del Cas9-sgARN).

La FIG. 7 ilustra un ejemplo de un NATNA de tipo cruzado con CRISPR de Clase 2 modificado en el que el extremo 5' de un Cpfl-crARN está unido covalentemente al extremo 3' de un Cas9-sgARN, y en el que ambos elementos ahorquillados 3' (por ejemplo, FIG. 2, 207, 208) se han eliminado del Cas9-sgARN. En la figura la flecha indica la región de unión entre el Cpfl-crARN y el Cas9-sgARN. El Cpfl-crARN y el Cas9-sgARN pueden unirse directamente entre ellos por un enlace covalente o, por ejemplo, unirse entre ellos mediante una secuencia de nucleótidos de elemento enlazador (por ejemplo, el extremo 5' de un Cpfl-crARN se une covalentemente al extremo 3' de una secuencia de nucleótido de elemento enlazador y el extremo 5' de la secuencia de nucleótidos del elemento enlazador se une covalentemente al extremo 3' de un Cas9-sgARN, en el que se han eliminado la segunda horquilla y la tercera horquilla del Cas9-sgARN).

Como se ha expuesto anteriormente, en algunas realizaciones, cada polinucleótido de cada una de los dos NATNA diferentes (véase, por ejemplo, la FIG. 1B, FIG. 2, y FIG. 3) comprende adicionalmente uno o más elementos enlazadores conectados covalentemente en el extremo 5' o el extremo 3'. Por ejemplo, un Cpfl-NATNA comprende adicionalmente uno o más elementos enlazadores conectados en el extremo 5' y/o el extremo 3', un primer Cas9-NATNA (por ejemplo, un Cas9-crARN) y/o un segundo Cas9-NATNA (por ejemplo, un Cas9-tracrARN) comprende adicionalmente uno o más elementos enlazadores en el extremo 5' y/o el extremo 3' del primer Cas9-NATNA y/o el segundo Cas9-NATNA, y un Cas9-NATNA sencillo (por ejemplo, un Cas9-sgARN) comprende adicionalmente uno o más elementos enlazadores conectados covalentemente en el extremo 5' y/o el extremo 3'.

En otras realizaciones, dos NATNA diferentes están conectados no covalentemente para formar un NATNA de tipo cruzado. Por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado puede comprender un primer Cas9-NATNA o un segundo Cas9-NATNA conectado no covalentemente con un Cpfl-NATNA mediante un enlace de hidrógeno de pares de bases en el extremo 5' o el extremo 3' del primer Cas9-NATNA, o el extremo 5' o el extremo 3' del segundo Cas9-NATNA. En realizaciones adicionales, para la conexión no covalente, los diferentes NATNA están conectados no covalentemente mediante una o más secuencias de elemento enlazador.

La FIG. 5A y FIG. 5B ilustran ejemplos de dichas uniones no covalentes utilizando ARN NATNA.

La FIG. 5A ilustra un ejemplo de un Cas9-Cpfl-NATNA modificado. En este ejemplo, el extremo 3' del Cas9-sgARN se modifica para eliminar un elemento ahorquillado 3' (FIG. 2, 207 o 208). La secuencia 1 del Cpfl-tallo de ARN unida covalentemente al Cas9-sgARN (FIG. 3, 303), incluyen la secuencia 5' del tallo Cpfl (aproximadamente de 3-6 nt), importante para la formación de un pseudo-nudo, está unido covalentemente al Cas9-sgARN que se modificó para retirar una horquilla 3'. El Cpfl-NATNA está modificado para retirar la secuencia 1 de Cpfl-tallo de ARN (FIG. 3, 303), elemento bucle (FIG. 3, 304) y la secuencia 5' del tallo Cpfl. La secuencia 1 del Cpfl-tallo de ARN (FIG. 3, 303) se puede hibridar mediante enlace de hidrógeno de pares de bases con la secuencia 2 Cpfl-crARN tallo del ARN (FIG. 3, 305). En la figura, la flecha indica la región de unión de hidrógeno de pares de bases entre la secuencia 2 de Cpfl-tallo ARN del Cpfl-crARN y Cas9-sgARN modificado para comprender la secuencia 1 de Cpfl-tallo de ARN, incluyendo la secuencia 5' del tallo Cpfl.

En otra realización, se pueden añadir secuencias adicionales al extremo 3' del Cas9-sgARN y/o se pueden añadir secuencias adicionales complementarias o no complementarias al extremo 5' del Cpfl-crARN. La FIG. 5B ilustra un ejemplo de dicho Cas9-Cpfl-NATNA modificado. En este ejemplo, el extremo 3' de un Cas9-sgARN se modifica por unión covalente de una primera secuencia de elemento enlazador y el extremo 5' de un Cpfl-crARN se modifica por unión covalente a una segunda secuencia de elemento enlazador, en el que el primer elemento enlazador y el segundo elemento enlazador se hibridan y forman una conexión mediante enlace de hidrógeno de pares de bases. En la figura, la flecha indica la región de unión de hidrógeno entre el primer elemento enlazador y el segundo elemento enlazador.

Además de los enlaces covalentes y enlaces no covalentes, se pueden utilizar otros tipos de conexiones para formar un NATNA de tipo cruzado entre el primer NATNA y el segundo NATNA incluyendo, pero sin limitarse a, emparejamientos de restos por unión ligando/ligando, y/o entrecruzamiento. Los emparejamientos de restos por unión ligando/ligando puede ser útil para formar un NATNA de tipo cruzado incluye, pero no se limita a: una secuencia de ácido nucleico seleccionada y su correspondiente aptámero, y una estructura secundaria de ácido nucleico/una molécula pequeña, ion, o proteína que se une a la estructura secundaria de ácido nucleico. Normalmente, un primer NATNA se adapta para comprender un ligando (por ejemplo, el primer NATNA comprende en su extremo 3' una

secuencia de ácido nucleico seleccionad) y un segundo NATNA se adapta para comprender un resto de unión al ligando (por ejemplo, el segundo NATNA comprende un aptámero en su extremo 5' que se une a la secuencia de ácido nucleico seleccionado).

5 Los agentes de entrecruzamiento útiles para formar un NATNA de tipo cruzado (por entrecruzamiento de un primer NATNA a un segundo NATNA) incluyen, pero no se limitan a: agentes alquilantes (por ejemplo, 1,3-bis(2-cloroetil-l-nitrosourea) y mostaza nitrogenada); cisplatino (cis-diaminodicloroplatino (II)) y sus derivados); radiación ionizante; ácido nitroso; reactivos químicos (por ejemplo, malondialdehído); psoralenos (activados en presencia de UV); y aldehídos (por ejemplo, acroleína y crotonaldehído).

10 En realizaciones preferidas, la presente invención incluye un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas (que comprenden un primer NATNA y un segundo NATNA) en el que el primer NATNA es capaz de formar un primer complejo con una primera proteína y el segundo NATNA es capaz de formar un segundo complejo con la segunda proteína. Además, si se forma el primer complejo, es capaz de dirigirse por unión específica del sitio a una segunda secuencia de ácido nucleico diana y, si se forma el segundo complejo, es capaz de dirigirse por unión específica del sitio a una segunda secuencia de ácido nucleico diana. En una realización, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, que comprende un CpfI-NATNA y un Cas9-NATNA, es capaz de formar un complejo CpfI-NATNA/proteína CpfI con una proteína CpfI y un complejo Cas9-NATNA/proteína Cas9 con una proteína Cas9 (se hace referencia al complejo nucleoproteico completo como un complejo CAS9-CpfI-NATNA/proteína Cas9&CpfI). Cuando el complejo Cas9-CpfI-NATNA/proteína CpfI se forma entre el Cas9-CpfI-NATNA y la proteína CpfI, el complejo es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble; y cuando el complejo Cas9-CpfI-NATNA/proteína Cas9 se forma entre el Cas9-NATNA (por ejemplo, el primer Cas9-NATNA y el segundo Cas9-NATNA o el Cas9-NATNA sencillo) y una proteína Cas9, el complejo Cas9-CpfI-NATNA/proteína Cas9 es capaz de unirse a una segunda secuencia de ácido nucleico diana de cadena doble. Por lo tanto, cuando el complejo Cas9-CpfI-NATNA/proteína Cas9&CpfI se forma, el complejo Cas9-CpfI-NATNA/proteína Cas9&CpfI es capaz de unirse a la primera secuencia de ácido nucleico diana de cadena doble y a la segunda secuencia de ácido nucleico diana de cadena doble.

30 El Ejemplo 1 describe la producción de componentes a modo de ejemplo de NATNA tipo cruzado modificados.

El Ejemplo 5 describe la identificación y exploración de crARN de Clase 2 que se pueden utilizar para fabricar los NATNA de tipo cruzado modificados de la presente invención. El Ejemplo 6 describe la identificación y exploración de tracrARN de Clase 2 que se pueden utilizar para fabricar los NATNA de tipo cruzado modificados de la presente invención.

35 El Ejemplo 8 describe un método de sondeo de los sitios en las estructuras de crARN guía de Clase 2 Tipo V que son tolerantes a la modificación. El Ejemplo 9 describe un método de sondas para los sitios tolerantes a la modificación en las estructuras de ARN guía de Clase 2 Tipo II Cas9. La información obtenida por los métodos del Ejemplo 8 y el Ejemplo 9 proporciona una guía respecto a los sitios de las estructuras mediante los cuales se pueden conectar los NATNA individuales para formar los NATNA de tipo cruzado modificados de la presente invención.

45 En un segundo aspecto, la presente invención se refiere a composiciones de ácido nucleico/proteína que comprenden un NATNA de tipo cruzado como se define en las reivindicaciones adjuntas (que comprenden un primer NATNA y un segundo NATNA), una primera proteína con la que el primer NATNA es capaz de formar un complejo, y una segunda proteína con la que el segundo NATNA es capaz de formar un complejo; por lo tanto, el NATNA de tipo cruzado, la primera proteína, y la segunda proteína son capaces de formar un complejo de NATNA de tipo cruzado/primer&segunda proteínas. Normalmente, la primera proteína comprende una o más actividades de nucleasa, y la segunda proteína comprende una o más actividades nucleasa. En algunas realizaciones, la primera proteína es catalíticamente inactiva para una o más de las actividades nucleasa, la segunda proteína es catalíticamente inactiva para una o más de las actividades nucleasa, o la primera proteína es catalíticamente inactiva para una o más de las actividades nucleasa y la segunda proteína es catalíticamente inactiva para una o más actividades nucleasa. Otras realizaciones del NATNA de tipo cruzado/primer&segunda proteínas, en las que la primera proteína o la segunda proteína es catalíticamente inactiva, incluyen la asociación de un polinucleótido donante con la proteína catalíticamente inactiva.

55 En una realización de este segundo aspecto de la presente invención, una composición de ácido nucleico/proteína comprende un NATNA de tipo cruzado con CRISPR de Clase 2 modificado como se describe en el presente documento y una proteína Cas9, una proteína CpfI, o ambas, una proteína Cas9 y una proteína CpfI. En otra realización, el NATNA de tipo cruzado con CRISPR de Clase 2 modificado está formando un complejo con la proteína Cas9 (complejo Cas9-CpfI-NATNA/proteína Cas9), la proteína CpfI (complejo Cas9-CpfI-NATNA/proteína CpfI), o ambas, la proteína Cas9 y la proteína CpfI (complejo Cas9-CpfI-NATNA/proteínas Cas9&CpfI). La proteína Cas9 y la proteína CpfI pueden tener combinaciones de las siguientes actividades endonucleasa: para la proteína Cas9, ambos dominios RuvC-1 y HNH de la proteína Cas9 pueden ser catalíticamente inactivos, el dominio RuvC-1 de la proteína Cas9 puede ser catalíticamente inactivo, el dominio HNH de la proteína Cas9 puede ser catalíticamente inactivo, y ambos dominios RuvC-1 y HNH pueden estar catalíticamente inactivos; y para la proteína CpfI, la proteína CpfI puede ser catalíticamente activa o catalíticamente inactiva.

En algunas realizaciones de la composición, cualquiera de entre la proteína Cas9 o la proteína Cpf1 es catalíticamente inactiva (dCas9 o dCpf1) y la composición comprende adicionalmente un polinucleótido donante en el que el polinucleótido donante comprende una secuencia de nucleótidos complementaria del elemento espaciador, o las regiones adyacentes al elemento espaciador del Cpf1-NATNA cuando está presente dCpf1, o una secuencia de nucleótidos complementaria al elemento espaciador, o las regiones adyacentes al elemento espaciador, del Cas9-NATNA cuando dCas9 está presente. El polinucleótido donante es capaz de asociarse con el elemento espaciador, o las regiones adyacentes al elemento espaciador, mediante unión de hidrogeno entre la secuencia de nucleótidos del polinucleótido donante complementaria al elemento espaciador, o la secuencia adyacente al elemento espaciador.

Las mutaciones de la proteína Cas9 que son enzimáticamente inactivas para la actividad nucleasa relacionada con RuvC-1, actividad nucleasa relacionada con HNH, y ambas, la actividad nucleasa relacionada con RuvC-1 y la actividad nucleasa relacionada con HNH se conocen en la técnica. Las mutaciones de la proteína Cpf1 que son enzimáticamente inactivas se conocen en la técnica (véase, por ejemplo, Yamano, T., et al., *Cell* 165(4):949-962 (2016)); Zetsche, B., et al, *Cell* 163:1-13 (2015)). Las proteínas Argonauta enzimáticamente inactivas se pueden producir por modificación de uno o ambos dominios nucleasa Argonauta (por ejemplo, los dominios KIWI y PAZ). Se puede modificar una proteína Argonauta deficiente en nucleasa mediante la modificación de secuencia de los restos catalíticos de los dominios de nucleasa, particularmente el dominio PIWI de la proteína Argonauta (Jinek, M., et al, *Nature* 457:405-412 (2009)). Adicionalmente, las proteínas Argonauta deficientes en nucleasa de origen natural se conocen en la técnica (por ejemplo, dominios PIWI inactivos), y se pueden utilizar como modelos para diseñar variantes inactivas de Argonautas activas en nucleasa (véase, por ejemplo, Makarova, K., et al, *Biology Direct* 4:29 (2009) [doi:10.1186/1745-6150-4-29]).

A través de los sistemas CRISPR, "la biogénesis del guía" (a la que se hace referencia también como "procesamiento del guía") implica el truncado de endonucleasa o exonucleasa de la secuencia de ARN guía a continuación de la transcripción de la matriz CRISPR. El procesamiento enzimático del ARN guía se puede llevar a cabo por RNasas codificadas por el operón Cas (por ejemplo, sistemas Cas 6 de Clase 1 Tipo I-E) o por RNasas endógenas (por ejemplo, sistemas de RNasas III de Clase 2 Tipo II-A).

En los sistemas Clase 2 Tipo V, la guía de biogénesis se lleva a cabo por la nucleasa proteica Cpf1. La proteína Cpf1 también es responsable de la escisión de ADN diana de cadena doble específica de la secuencia.

En el sistema Tipo V, la escisión del pre-crARN (véase, por ejemplo, la FIG. 14B) se produce en una región corriente arriba (por ejemplo, en la dirección 5') desde la estructura secundaria pseudo-nudo y da como resultado la generación de un Cpf1 crARN guía (véase, por ejemplo, la FIG. 14A). En algunas realizaciones de la presente invención, es útil evitar que la proteína Cpf1 se escinda 5' del elemento tallo del crARN guía, por ejemplo, para evitar la separación de un complejo Cas9-Cpf1-NATNA/proteínas Cas9&Cpf1 en un complejo Cas9-NATNA/proteína Cas9 y un complejo Cpf1-NATNA /proteína Cpf1 que se produce como resultado de la escisión de la proteína Cpf1 del complejo Cpf1-NATNA/proteínas Cas9&Cpf1. Se ha demostrado que la secuencia del pre-crARN Tipo V se puede modificar para evitar el procesamiento del ARN guía por la proteína Cpf1 CRISPR Tipo V (véase, Fonfara, I., et al, *Nature* 532(7600):517-521 (2016)).

Un método para evitar la escisión de Cpf1 de secuencias 5' del elemento tallo del crARN guía es por la modificación (por ejemplo, mutaciones, inserciones, eliminaciones o modificaciones químicas de bases) de las bases de la región corriente arriba del pseudo-nudo o en el pseudo-nudo del pre-ARN para evitar el procesamiento del pre-crARN por la proteína Cpf1. Para evaluar el efecto de dichas modificaciones sobre el procesamiento del guía, el pre-crARN modificado se incuba en presencia de un equivalente de la proteína Cpf1 durante un periodo de tiempo en un tampón adecuado. La mezcla se trata con Proteinasa K (Denville Scientific, South Plainfield, NJ) para retirar la proteína y la mezcla se analiza por electroforesis en gel de poliacrilamida para evaluar si se produce la escisión del pre-crARN modificado. Un pre-crARN que no se incuba en presencia de un equivalente de la proteína Cpf1 sirve como control positivo (es decir, un control de la ausencia de procesamiento del guía). Si no es suficiente una modificación sencilla en el pre-crARN para anular el procesamiento del guía, entonces se pueden combinar las combinaciones de modificaciones que presentan un procesamiento reducido del pre-crARN en un diseño de pre-crARN y se reensaya en cuanto a la ausencia de actividad de procesamiento del guía. Las modificaciones del pre-ARN que dan como resultado la incapacidad del pre-ARN modificado para procesarse se puede evaluar adicionalmente en cuanto a la capacidad del complejo Cpf1-pre-crARN/proteína Cpf1 para mantener la unión específica de secuencia y/o escisión de un ácido nucleico ADN diana que comprende un elemento espaciador del pre-crARN.

Un segundo método para evitar la escisión de Cpf1 de secuencias 5' del elemento tallo del crARN guía es por la modificación de la proteína Cpf1. En este método, los restos de aminoácidos de la proteína Cpf1 se modifican para perturbar el procesamiento del guía. Se ha demostrado por cristalografía de rayos X del crARN guía/proteína Cpf1 que el pseudo-nudo se une por la interfaz de dos dominios proteicos denominados el dominio cuña (WED) y el dominio RuvC (véase Yamano, T., et al, *Cell* 165(4):949-962 (2016)). Los restos de aminoácido del Cpf1 proximales a la región de unión del extremo 5' del crARN guía y/o la estructura pseudo-nudo es probable que estén implicados en la catálisis por endonucleasa de los pre-crARN. Se pueden utilizar estrategias de mutagénesis, tal como exploración de alanina (véase, por ejemplo, Lefevre, F., et al, *Nucleic Acids Research* 25(2):447-448 (1997); Lee, et al, *Molecular*

- Pharmacology 50(1): 140-148 (1996)) para modificar regiones en el dominio WED y RuvC, u otros dominios en la proteína CpfI, para identificar restos en la proteína responsables del procesamiento del crARN guía. En este método, las proteínas CpfI que comprenden mutaciones de alanina se pueden expresar e incubar con un pre-crARN equivalente en un tampón adecuado. Después de la incubación, se añade Proteinasa K a la mezcla de reacción para eliminar la
- 5 proteína CpfI y la mezcla de reacción se analiza entonces mediante electroforesis en gel de poliacrilamida para evaluar si se produjo la escisión del pre-crARN modificado. Un pre-crARN no incubado en presencia de una proteína CpfI equivalente sirve como control positivo (es decir, un control de la ausencia del procesamiento del guía. Si la mutación sencilla en la proteína CpfI no es suficiente para anular el procesamiento del guía, entonces se pueden combinar en una única construcción de proteína CpfI combinaciones de mutaciones que presenten un procesamiento reducido del
- 10 pre-crARN y re-ensayarse en cuanto a la ausencia de actividad de procesamiento del guía. Las mutaciones o combinaciones de mutaciones de la proteína CpfI se pueden evaluar adicionalmente en cuanto a la capacidad del complejo CpfI-pre-crARN para mantener la unión específica de secuencia y/o escisión de un ácido nucleico ADN diana que comprende el elemento espaciador pre-crARN.
- 15 En un tercer aspecto, la presente invención se refiere a secuencias de ácido nucleico que codifican un NATNA de tipo cruzado (que comprende un primer NATNA y un segundo NATNA), así como casetes de expresión, vectores y células recombinantes que comprende secuencias de ácido nucleico que codifican NATNA de tipo cruzado como se define por las reivindicaciones adjuntas. Una realización del tercer aspecto de la presente invención se refiere a una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado. En algunas
- 20 realizaciones del tercer aspecto de la invención, dichos casetes de expresión, vectores y células recombinantes comprenden además secuencias que codifican una primera proteína (por ejemplo, una proteína CpfI) que la cual el primer NATNA es capaz de formar un complejo, y/o una segunda proteína (por ejemplo, una proteína Cas9) con la que el segundo NATNA es capaz de formar un complejo.
- 25 En una realización, la presente invención se refiere a uno o más casetes de expresión que comprende una o más secuencias de ácido nucleico que codifica un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas, y opcionalmente una o más secuencias de ácido nucleico que codifican una primera proteína con la que un NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias de ácido
- 30 nucleico que codifican una proteína CpfI y/o una proteínas Cas9). Los casetes de expresión comprenden normalmente secuencias reguladoras que están implicadas en una o más de las siguientes: regulación de la transcripción, regulación post-transcripcional, y regulación de la traducción. Los casetes de expresión se pueden introducir en una amplia variedad de organismos que incluyen células bacterianas, células de levadura, células vegetales y células de mamífero. Los casetes de expresión normalmente comprenden secuencias reguladoras funcionales que corresponden a los organismos en los que se introducen.
- 35 En algunas realizaciones, una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas, y opcionalmente una o más secuencias de ácido nucleico que codifican una primera proteína con la cual el NATNA tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la cual el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, opcionalmente una o más secuencias de ácido nucleico que codifica una proteína CpfI y/o una proteína Cas9) están unidos operativamente a los elementos reguladores.
- 40 Una realización adicional de la presente invención se refiere a vectores, que incluyen vectores de expresión, que comprenden una o más secuencias de ácidos nucleicos que codifican un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas, y opcionalmente una o más secuencias de ácido nucleico que codifican una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, una o más secuencias de ácido nucleico que codifica un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias de ácido nucleico que codifican una proteína CpfI y/o una proteína Cas9). Los vectores pueden incluir también secuencias que codifican marcadores genéticos o explorables. Además, también se pueden añadir secuencias de direccionamiento nuclear, por ejemplo, a las secuencias codificantes de la proteína Cas9 y la proteína CpfI. Los vectores también pueden incluir polinucleótidos que codifican marcadores proteicos (por ejemplo, marcadores poli-His, marcadores de hemaglutinina, marcadores de proteínas fluorescentes, marcadores bioluminiscentes). Las secuencias codificantes de dichos marcadores proteicos se pueden fusionar, por ejemplo, la una o más secuencias de aminoácidos que codifican una proteína Cas9 y/o una proteína CpfI.
- 45 Se conocen métodos generales para la construcción de vectores de expresión en la técnica. Los vectores de expresión para las células huésped están disponibles en el mercado. Existen varios productos de software comerciales diseñados para facilitar la selección de vectores apropiados y la construcción de los mismos, tales como vectores de células de insectos para la transformación de células de insecto y la expresión en células de insecto, plásmidos bacterianos para la transformación bacteriana y la expresión genética en células bacterianas, plásmidos de levadura para la transformación celular y la expresión genética en levaduras y otros hongos, vectores de mamífero para la transformación en células de mamífero y la expresión genética en células de mamífero o mamíferos, y vectores víricos (que incluyen vectores lentivirus, retrovirus, adenovirus, virus del herpes simples I o II, parvovirus, virus de
- 50  
55  
60  
65

reticuloendoteliosis, y virus adeno-asociados (AAV)) para la transformación y expresión genética y métodos para permitir fácilmente la clonación de dichos polinucleótidos. Los vectores ilustrativos de la transformación en plantas incluyen los derivados de un plásmido Ti de *Agrobacterium tumefaciens* (Lee, L.Y., et al, Plant Physiology 146(2): 325-332 (2008)). También son útiles y se conocen en la técnica los plásmidos de *Agrobacterium rhizogenes*. Por ejemplo, el SNAPGENE™ (GSL Biotech LLC, Chicago, IL; [snappgene.com/resources/plasmid\\_files/yourtimeisvaluable/](http://snappgene.com/resources/plasmid_files/yourtimeisvaluable/)) proporciona una extensa lista de vectores, secuencias de vectores individuales, y mapas de vectores, así como fuentes comerciales para muchos de los vectores.

Los vectores lentivíricos son ejemplos de vectores útiles para la introducción en células de mamífero de una o más secuencias de ácido nucleico y NATNA de tipo cruzado modificado, y opcionalmente una o más secuencias de ácido nucleico que codifican una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la cual el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, una o más secuencias que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias que codifican una proteína Cpf1 y/o una proteína Cas). Los lentivirus son miembros de la familia Retroviridae y es un virus ARN de cadena sencilla, que puede infectar células tanto en división como en no división, así como proporcionan una expresión estable mediante la integración en el genoma. Para aumentar la seguridad del lentivirus, se dividen los componentes necesarios para producir un vector vírico a través de múltiples plásmidos. Los vectores de transferencia son normalmente incompetentes para la replicación y pueden contener adicionalmente una eliminación en el 3' LTR, lo que da lugar a que el virus se auto-inactiva después de la integración. El empaquetamiento y envuelta de plásmidos se utiliza normalmente en combinación con un vector de transferencia. Por ejemplo, un plásmido empaquetado puede codificar combinaciones de los genes Gag, Pol, Rev, y Tat. Un plásmido de transferencia puede comprender LTR víricos y la señal de empaquetamiento psi. La envoltura de los plásmidos comprende una proteína de envoltura (habitualmente la glicoproteína del virus de estomatitis vesicular, VSV-GP debido a su amplio intervalo de infectividad).

Los vectores lentivíricos basados en el virus de inmunodeficiencia humana Tipo-1 (VIH-1) tiene proteínas accesorias adicionales que facilitan la integración en ausencia de división celular. Los vectores VIH-1 se han diseñado para afrontar varios problemas de seguridad. Estos incluyen la expresión por separado de los genes víricos en trans para evitar los casos de recombinación que dan lugar a la generación de virus competentes en replicación. Además, el desarrollo de vectores auto-inactivantes reduce el potencial de transactivación de los genes de la vecindad y permite la incorporación de elementos reguladores para dirigir la expresión genética en ciertos tipos celulares particulares (véase, por ejemplo, Cooray, S., et al, Methods in Enzymology 507:29-57 (2012)).

Las células huésped transformadas (o células recombinantes) son células o la progenie de células que se han transformado o transfectado utilizando técnicas de ADN recombinante, con una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado modificado, y opcionalmente una o más secuencias de ácido nucleico que codifican una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias de ácido nucleico que codifican una proteína Cpf1 y/o una proteína Cas9). Los métodos de introducción de polinucleótidos (por ejemplo, en un vector de expresión) en células huésped se conocen en la técnica y se seleccionan normalmente basándose en la clase de célula huésped. Dichos métodos incluyen, por ejemplo, infección vírica o por bacteriófagos, transfección, conjugación, electroporación, precipitación en fosfato cálcico, transfección mediada por polietilenglicol, transfección mediada por DEAE-dextrano, fusión de protoplastos, lipofección, transfección mediada por liposomas, tecnología de pistola de partículas, microinyección directa y suministro mediado por nanopartículas.

Como alternativa para expresar una o más secuencias de ácido nucleico que codifiquen un NATNA de tipo cruzado modificado, y opcionalmente una o más secuencias de ácido nucleico que codifiquen una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, se pueden introducir un NATNA de tipo cruzado y/o la primera y segunda proteínas en una células, por ejemplo. O uno o más componentes se pueden expresar por la célula y el otro componente se introduce directamente. Los métodos para la introducción de los componentes en una célula incluyen electroporación, lipofección, y transferencia genética balística (por ejemplo, utilizando una pistola genética o un sistema de suministro de partículas biolístico).

Se desvelan en el presente documento varias células huésped a modo de ejemplo que se pueden utilizar para producir células recombinantes por la introducción de una o más secuencias que codifican un NATNA de tipo cruzado modificado, y opcionalmente una o más secuencias que codifican una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y/o una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo (por ejemplo, una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias de ácido nucleico que codifican una proteína Cpf1 y/o una proteína Cas9). Dichas células huésped incluyen, pero no se limitan a una célula vegetal, una célula de levadura, una célula bacteriana, una célula de insecto, una célula de alga, o una célula de mamífero.

Los métodos de introducción de polinucleótidos (por ejemplo, un vector de expresión) en células huésped para producir

células recombinantes se conocen en la técnica y se seleccionan normalmente basándose en la clase de célula huésped. Dichos métodos incluyen, por ejemplo, infección vírica o por bacteriófagos, transfección, conjugación, electroporación, precipitación en fosfato cálcico, transfección mediada por polietilimina, transfección mediada por DEAE-dextrano, fusión de protoplastos, lipofección, transfección mediada por liposomas, tecnología de pistola de partículas, microinyección directa y suministro mediado por nanopartículas. Por facilidad de exposición, "transfección" se utiliza posteriormente para referirse a cualquier método de introducción de polinucleótidos en una célula huésped.

Los métodos preferidos para la introducción de polinucleótidos en células vegetales incluyen el bombardeo de microproyectiles y la transformación mediada por *Agrobacterium*. De manera alternativa, se pueden utilizar otras especies no *Agrobacterium* (por ejemplo, *Rhizobium*) y otras células procariontas que sean capaces de infectar células vegetales e introducir polinucleótidos heterólogos en el genoma de la célula vegetal infectada. Otros métodos incluyen electroporación, transfección mediada por liposomas, transformación utilizando polen o virus, y agentes químicos que aumentan la captación de ADN libre, o el suministro de ADN libre utilizando bombardeo de microproyectiles. Véase, por ejemplo, Narusaka, Y., et al, Capítulo 9, en *Transgenic Plants - Advances and Limitations*, editado por Yelda, O., ISBN 978-953-51-0181-9 (2012).

En algunas realizaciones, una célula huésped se transfecta transitoria o no transitoriamente. En algunas realizaciones, una célula se transfecta como se produce naturalmente en un sujeto. En algunas realizaciones, una célula que se va a transfectar se toma de un sujeto, por ejemplo, una célula primaria o célula progenitora. En algunas realizaciones, la célula primaria o célula progenitora se cultiva y/o se devuelve después de la transfección *ex vivo* al mismo sujeto (tratamiento autólogo) o a un sujeto diferente.

Los complejos NATNA de tipo cruzado/primerasegunda proteínas descritos en el presente documento se pueden utilizar para generar organismos transgénicos no humanos por introducción específica del sitio, de una secuencia de polinucleótido seleccionada en un locus de ADN diana en el genoma para generar una modificación del ADN genómico. El organismo transgénico puede ser un animal o un vegetal.

un animal transgénico se genera normalmente introduciendo el sistema en una célula del cigoto. Una técnica básica, descrita en referencia a la fabricación de ratones transgénicos (Cho, A., et al, "Generation of Transgenic Mice," *Current Protocols in Cell Biology*, CAPÍTULO. Unit-19.11 (2009)), implica cinco pasos básicos: primero, preparación de un sistema, como se describe en el presente documento, que incluye un polinucleótido donante adecuado; segundo, recolectar los cigotos donantes; tercero, microinyección del sistema en el cigoto de ratón; cuarto, implantación de los cigotos microinyectados en ratones receptores pseudo-gestantes; y quinto, realización del genotipo y análisis de la modificación del ADN genómico establecido en los ratones fundadores. Los ratones fundadores pasarán la modificación a cualquier progenie. Los ratones fundadores son normalmente heterocigóticos para el transgén. El cruzamiento entre estos ratones producirá ratones que serán homocigóticos para el transgén el 25 % de las veces.

Los métodos para la generación de plantas transgénicas también son bien conocidos. Una planta transgénica generada, por ejemplo, utilizando métodos de transformación con *Agrobacterium*, contiene normalmente un transgén insertado en un cromosoma. Es posible producir una planta transgénica que sea homocigótica con respecto a un transgén reproduciendo sexualmente (es decir, por autofecundación) una planta transgénica segregante independiente que contiene un único transgén con ella misma, por ejemplo, una planta F0, para producir una semilla F1. Las plantas formadas por la germinación de las semillas F1 se pueden ensayar en cuanto a homocigosis. Los ensayos de cigosis típicos incluyen, pero no se limitan a ensayos de polimorfismo de un único nucleótido y ensayos de amplificación térmica que distinguen entre homocigotos y heterocigotos.

En un cuarto aspecto, se desvelan métodos de utilización de composiciones de ácido nucleico/proteínas que comprende un NATNA de tipo cruzado modificado, una primera proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo, y una segunda proteína con la que el NATNA de tipo cruzado es capaz de formar un complejo. Se describen en el presente documento aspectos de composiciones de ácido nucleico/proteína, por ejemplo, en el segundo aspecto anterior.

En una realización, la presente invención incluye un método de unión de una secuencia de ácido nucleico (por ejemplo, ADN), que comprende poner en contacto la primera secuencia de ácido nucleico diana del ácido nucleico (por ejemplo, ADN) y una segunda secuencia de ácido nucleico diana en la secuencia de ácido nucleico (por ejemplo, ADN) con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas formando un complejo con una proteína con la que el NATNA de tipo cruzado forma un complejo, y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpf1 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), de manera que se facilite la unión de la composición de ácido nucleico/proteína con la primera secuencia de ácido nucleico diana de la secuencia de ácido nucleico y la segunda secuencia de ácido nucleico diana del ácido nucleico, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El NATNA de tipo cruzado modificado comprende un primer elemento espaciador de NATNA (por ejemplo, un elemento espaciador Cpf1-NATNA) que es complementario de la primera secuencia de ácido nucleico diana (por ejemplo, ADN) y un segundo espaciador de NATNA (por ejemplo, un elemento espaciador Cas9-NATNA) que es

complementario de la segunda secuencia de ácido nucleico diana (por ejemplo, ADN). En algunas realizaciones, la secuencia de ácido nucleico diana es ADN o ADN genómico. Dichos métodos de unión de una secuencia de ácido nucleico diana se llevan a cabo, *in vitro*, en una célula (por ejemplo, en células cultivadas), *ex vivo* (por ejemplo, en células madre retiradas de un sujeto), e *in vivo*.

5 Se conocen en la técnica varios métodos para evaluar y/o cuantificar las interacciones proteína-ácido nucleico que incluyen, pero no se limitan a, los siguientes: ensayos de inmunoprecipitación (ChIP), ensayos de cambio de movilidad electroforética de ADN (EMSA), ensayos de despliegue de ADN, y ensayos de captura y detección en microplaca. Están disponibles kits, materiales y reactivos en el comercio para la práctica de muchos de estos métodos en, por ejemplo, Thermo Scientific (Wilmington, DE), Signosis (Santa Clara, CA), Bio-Rad (Hercules, CA), y Promega (Madison, WI)). Una estrategia común para detectar las interacciones proteína-ácido nucleico es EMSA (véase, por ejemplo, Hellman L.M., et al, Nature Protocols 2(8):1849—1861 (2007)).

15 En otra realización, la presente invención incluye un método de unión de una secuencia de ácido nucleico (por ejemplo, ADN), que comprende poner en contacto la primera secuencia de ácido nucleico diana del ácido nucleico (por ejemplo, ADN) y una segunda secuencia de ácido nucleico diana en la secuencia de ácido nucleico (por ejemplo, ADN) con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas formando un complejo con una proteína con la que el NATNA de tipo cruzado forma un complejo, y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpf1 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), de manera que se facilite la unión de la composición de ácido nucleico/proteína con la primera secuencia de ácido nucleico diana de la secuencia de ácido nucleico y la segunda secuencia de ácido nucleico diana del ácido nucleico, en el que el método no comprende un procedimiento para la modificación de la identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El NATNA de tipo cruzado modificado comprende un primer elemento espaciador de NATNA (por ejemplo, un elemento espaciador Cpf1-NATNA) que es complementario de la primera secuencia de ácido nucleico diana (por ejemplo, ADN) y un segundo elemento espaciador de NATNA (por ejemplo, un elemento espaciador Cas9-NATNA) que es complementario de la segunda secuencia de ácido nucleico diana (por ejemplo, ADN). La primera proteína (por ejemplo, Cpf1) de la composición de ácido nucleico/proteína unida corta la primera secuencia de ácido nucleico diana, y la segunda proteína (por ejemplo, Cas9) de la composición de ácido nucleico/proteína unida corta la segunda secuencia de ácido nucleico diana. En algunas realizaciones, la secuencia de ácido nucleico diana es ADN o ADN genómico. Dichos métodos de unión de una secuencia de ácido nucleico diana se llevan a cabo, *in vitro*, en una célula (por ejemplo, en células cultivadas), *ex vivo* (por ejemplo, en células madre retiradas de un sujeto), e *in vivo*.

El Ejemplo 3 describe la ejecución de los ensayos de escisión mediada por proteína Cas. El Ejemplo 2 describe como producir regiones de ADN diana de cadena doble para su uso en los ensayos de escisión.

40 El Ejemplo 4 presenta un análisis de secuenciación profundo para la detección de modificaciones diana en células eucariotas utilizando los NATNA de tipo cruzado de la presente invención. El ejemplo 7 presenta un análisis alternativo, el ensayo T7E1, para la detección de modificaciones diana en las células eucariotas utilizando los NATNA de tipo cruzado de la presente invención.

45 El Ejemplo 10 ilustra el uso de los NATNA de tipo cruzado de la presente invención para seleccionar y modificar secuencias de ADN diana presentes en el ADN genómico y para medir el nivel de actividad de escisión en esos sitios.

Un método de corte de una secuencia diana de ácido nucleico se ejemplifica en la FIG. 11A, FIG. 11B, FIG. 11C, y FIG. 11D. Estas figuras presentan un ejemplo de escisión de una secuencia de una diana de ácido nucleico utilizando un Cas9-Cpf1-NATNA modificado y las proteínas Cas9 y Cpf1 teniendo ambas endonucleasas activas. La FIG. 11A muestra una proteína Cas9 (FIG. 11A, 1101) y una proteína Cpf1 (FIG. 11A, 1102), un Cas9-Cpf1-NATNA modificado como se muestra en la FIG. 4 (FIG. 11A, 1103), y un ácido nucleico de cadena doble (FIG. 11A, 1104) que comprende una primera secuencia de unión a la diana ADN complementaria del elemento espaciador Cas-sgARN (FIG. 11A, 1105) y una segunda secuencia de unión a la diana ADN complementaria del elemento espaciador Cpf1-crARN (FIG. 11A, 1106). La FIG. 11B ilustra las proteínas Cas9 (FIG. 11B, 1101) y Cpf1 (fig. 11B, 1102) en un complejo con Cas9-Cpf1-NATNA modificado (FIG. 11B, 1103) y la unión de hidrógeno de una primera secuencia de unión del ADN diana al elemento espaciador Cas9-sgARN (FIG. 11B, 1107) y una segunda secuencia de unión del ADN diana al elemento espaciador Cpf1-crARN (FIG. 11B, 1109). En la figura, el ácido nucleico de cadena doble se indica como FIG. 11B, 1194, la secuencia Cas9 PAM en el ácido nucleico de cadena doble se indica como FIG. 11B, 1108 y la secuencia Cpf1 PAM se indica como FIG. 11B, 1110. La FIG. 11C ilustra los cortes de los extremos romos de cadena doble producidos por Cas9 en la primera secuencia de unión del ADN diana (FIG. 11C, 1111) y los cortes escalonados de cadena doble producidos por Cpf1 en la segunda secuencia de unión del ADN diana (FIG. 11C, 1112). El ácido nucleico de cadena doble se indica como FIG. 11C, 1104. La FIG. 11D ilustra una secuencia de ácido nucleico escindido (FIG. 11D, 1113) que aún está asociado con el complejo Cas9-Cpf1-NATNA modificado (FIG. 11D, 1114). Un extremo del ácido nucleico diana comprende una rotura escalonada de cadena doble que resulta de la escisión con Cpf1 (FIG. 11D, 1112) y el otro una rotura del extremo de cadena doble esencialmente roma que resulta de la escisión con Cas9 (FIG.

11D, 1111). El ácido nucleico de cadena doble se indica como FIG. 11D, 1104. En otras realizaciones, las proteínas Cas9 y Cpf1 son deficientes en actividad endonucleasa y la unión del complejo Cas9-Cpf1-NATNA modificado a un ácido nucleico diana se puede utilizar, por ejemplo, para bloquear la transcripción y la expresión posterior de un gen adyacente al ácido nucleico diana.

5 En otra realización más, la presente invención incluye un método de modificación del ADN de una célula, que comprende poner en contacto una primera secuencia de ADN diana del ADN y una segunda secuencia de ADN diana del ADN con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas en un complejo con una primera proteína con la que el NATNA de tipo cruzado forma un complejo y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpf1 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia de ácido nucleico diana en la secuencia de ácido nucleico y la segunda secuencia de ácido nucleico diana del ácido nucleico, en el que el método no comprende un procedimiento para modificar la identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El NATNA de tipo cruzado modificado comprende un primer elemento espaciador NATNA (por ejemplo, un elemento espaciador Cpf1-NATNA) que es complementario de la primera secuencia de ADN diana y un segundo elemento espaciador NATNA (por ejemplo, un elemento espaciador Cas9-NATNA) que es complementario de la segunda secuencia de ADN diana. (por ejemplo, un ADN). La primera proteína (por ejemplo, Cpf1) de la composición unida de ácido nucleico/proteína corta la primera secuencia de ADN diana, y la segunda proteína (por ejemplo, Cas9) de la composición unida de ácido nucleico/proteína corta la segunda secuencia de ADN diana. La célula repara el primer sitio de corte y el segundo sitio de corte. Las rutas de reparación de ADN celulares incluyen, HDR, NHEJ, MMEJ. En algunas realizaciones, la secuencia de ácido nucleico diana es ADN o ADN genómico. Dichos métodos de unión de una secuencia de ácido nucleico diana se llevan a cabo *in vitro*, en la célula (por ejemplo, en células cultivadas), *ex vivo* (por ejemplo, en células madre retiradas de un sujeto), e *in vivo*. La etapa de contracción puede comprender adicionalmente que esté presente un polinucleótido donante, en el que al menos una parte del polinucleótido donante se incorpora entre el primer sitio de corte y el segundo sitio de corte.

30 En otra realización, la invención se refiere a un método para poner un polinucleótido donante en proximidad de una rotura de cadena doble en un ácido nucleico diana, normalmente un ADN, en una célula, en el que el método no comprende un procedimiento para la modificación de la identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El método comprende poner en contacto una primera secuencia de ADN diana del ADN y una segunda secuencia de ADN diana en un polinucleótido donante con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas en un complejo con una primera proteína con la que el NATNA de tipo cruzado forma un complejo y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpf1 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia del ADN diana del ADN y la segunda secuencia de ADN diana del polinucleótido donante. La primera secuencia de ADN diana es complementaria de un primer espaciador NATNA (por ejemplo, un elemento espaciador Cpf1-NATNA o un elemento espaciador Cas9-NATNA), en el que la primera proteína asociada es una nucleasa proteica catalíticamente activa (por ejemplo, una proteína Cpf1 o una proteína Cas9, respectivamente). La segunda secuencia de ADN diana es complementaria de un segundo espaciador NATNA (por ejemplo, un elemento espaciador Cpf1-NATNA o un elemento espaciador Cas9-NATNA), en el que la segunda proteína asociada es una nucleasa proteica catalíticamente inactiva (por ejemplo, una proteína dCpf1 o una proteína dCas9, respectivamente). Por ejemplo, cuando la nucleasa proteica catalíticamente activa es una proteína Cpf1, la nucleasa proteica catalíticamente inactiva es una proteína dCas9, y viceversa. La nucleasa proteica catalíticamente activa de la composición unida de ácido nucleico/proteína corta la primera secuencia de ADN diana para formar un sitio de corte. El polinucleótido donante está en proximidad al sitio de corte (por ejemplo, la rotura de cadena doble) debido a la nucleasa proteica catalítica activa y la nucleasa proteica catalíticamente inactiva están formando un complejo con el NATNA de tipo cruzado, es decir, forman parte de la misma composición de ácido nucleico/proteína. En algunas realizaciones, al menos una parte del polinucleótido donante se introduce en el sitio de corte del ADN (por ejemplo, mediante un proceso de reparación HDR) dando como resultado la modificación del ADN.

Las FIG. 12A, FIG. 12B y FIG. 12C ilustran la utilización de un Cas9-Cpf1-NATNA modificado y proteínas Cas9 y Cpf1, en el que los dominios de endonucleasa de Cas9 son activos y el dominio endonucleasa de Cpf1 es inactivo, para poner un polinucleótido donante en proximidad de una rotura de cadena doble en un ácido nucleico diana. La FIG. 12A muestra una proteína Cas9 (FIG. 12A, 1201) y una proteína Cpf1 (FIG. 12A, 1202), un Cas9-Cpf1-NATNA modificado como se muestra en la FIG. 4 (FIG. 12A, 1203), y un ácido nucleico de cadena doble (FIG. 12A, 1204) que comprende una primera secuencia de unión del ADN diana complementaria del elemento espaciador Cpf1-crARN (FIG. 12A, 1206). La FIG. 12B ilustra las proteínas Cas 9 (FIG. 12B, 1201) y Cpf1 (FIG. 12B, 1202) en un complejo con el Cas9-Cpf1-NATNA modificado (FIG. 12B, 1203), y la unión de hidrógeno de la primera secuencia de unión del ADN diana al elemento espaciador Cas9-sgARN (FIG. 12B, 1208) y la segunda secuencia de unión del ADN diana del polinucleótido donante al elemento espaciador Cpf1-crARN (FIG. 12B, 1210). En la figura, la secuencia Cas9 PAM del

ácido nucleico de cadena doble se indica como FIG. 12B, 1209 y la secuencia Cpfl PAM del polinucleótido donante se indica como FIG. 12B, 1211. El ácido nucleico de cadena doble se indica como FIG. 12B, 120, y el polinucleótido donante se indica como FIG. 12B, 1207. La FIG. 12C ilustra los cortes de los extremos truncados de cadena doble producidos por la Cas9 en la primera secuencia de unión del ADN diana (FIG. 12C, 1212) y muestra el polinucleótido donante (FIG. 12C, 1207) en proximidad de los cortes de los extremos truncados de cadena doble. Tener el polinucleótido donante (FIG. 12C, 1207) en estrecha proximidad a los cortes de cadena doble (FIG. 12C, 1212) aumenta la probabilidad de integración de las secuencias del polinucleótido donante, o partes de las mismas, en el ácido nucleico de cadena doble que comprende el ácido nucleico diana (FIG. 12C, 1204). En algunas realizaciones, el dominio endonucleasa de Cpfl es activo y capaz de unirse y escindir la molécula donante, pero Cpfl no se disocia del donante después de la escisión.

Las FIG. 13A, FIG. 13B y FIG. 13C ilustran la utilización de un Cas9-Cpfl-NATNA modificado y las proteínas Cas9 y Cpfl, en el que los dominios endonucleasa de Cas9 están inactivos y el dominio endonucleasa de Cpfl está activo, para poner un polinucleótido donante en proximidad de una rotura de cadena doble en un ácido nucleico diana. La FIG. 13A muestra una proteína Cas9 (FIG. 13A, 1301) y una proteína Cpfl (FIG. 13A, 1302), un Cas9-Cpfl-NATNA modificado como se muestra en la FIG. 4 (FIG. 13A, 1303), y un ácido nucleico de cadena doble (FIG. 13A, 1304) que comprende una primera secuencia de unión del ADN diana complementaria del elemento espaciador Cpfl-crARN (FIG. 13A, 1306) y un polinucleótido donante (FIG. 13A, 1307) que comprende una segunda secuencia de unión del ADN diana del elemento espaciador Cas9-sgARN (FIG. 13A, 1305). La FIG. 13B ilustra las proteínas Cas9 (FIG. 13B, 1301) y Cpfl (FIG. 13B, 1302) en complejo con el Cas9-Cpfl-NATNA modificado (FIG. 13B 1303) y el enlace de hidrógeno de la primera secuencia de unión del ADN diana con el elemento espaciador Cpfl-crARN (FIG. 13B, 1308) y la segunda secuencia de unión del ADN diana del polinucleótido donante al elemento espaciador Cas9-sgARN (FIG. 13B, 1310). En la FIG. 13B, la secuencia Cpfl PAM del ácido nucleico de cadena doble se indica como 1309 y la secuencia Cas9 PAM del polinucleótido donante se indica como 1311. El ácido nucleico de cadena doble que comprende la primera secuencia de unión del ADN diana se indica como FIG. 13B, 1304. La FIG. 13C ilustra cortes escalonados de cadena doble producidos por Cpfl en la primera secuencia de unión del ADN diana (FIG. 13C, 1312) y muestra el polinucleótido donante (FIG. 13C, 1307) en proximidad a los cortes escalonados de cadena doble. Tener el polinucleótido donante (FIG. 13C, 1307) en estrecha proximidad a los cortes de cadena doble (FIG. 13C, 1312) aumenta la probabilidad de integración de las secuencias de polinucleótido donante, o partes de las mismas, en el ácido nucleico de cadena doble que comprende la primera secuencia de unión del ADN diana (FIG. 13C, 1304). En algunas realizaciones, los dominios endonucleasa de Cas9 están activos y son capaces de unirse y escindir la molécula donante, pero Cas9 no se disocia del donante después de la escisión.

En una realización adicional, la invención se refiere a un método para poner un primer sitio de ácido nucleico diana, normalmente ADN, en proximidad de un segundo sitio de un ácido nucleico diana, normalmente un ADN, en una célula, en el que el método no comprende un procedimiento para la modificación de la identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El método comprende poner en contacto una primera secuencia de ácido nucleico diana y una segunda secuencia de ácido nucleico diana con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas en un complejo con una primera proteína con la que el NATNA de tipo cruzado forma un complejo y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpfl con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia del ácido nucleico diana y la segunda secuencia de ácido nucleico diana. La primera secuencia de ADN diana es complementaria de un primer espaciador NATNA (por ejemplo, un elemento espaciador Cpfl-NATNA o un elemento espaciador Cas9-NATNA), en el que la primera proteína asociada es un nucleasa proteica catalíticamente inactiva (por ejemplo, una proteína dCpfl o una proteína dCas9, respectivamente). La segunda secuencia de ADN diana es complementaria de un segundo espaciador NATNA (por ejemplo, un elemento espaciador Cpfl-NATNA o un elemento espaciador Cas9-NATNA), en el que la segunda proteína asociada es una nucleasa proteica catalíticamente inactiva (por ejemplo, una proteína dCpfl o una proteína dCas9, respectivamente). Por ejemplo, si la primera proteína es una proteína dCpfl, la segunda proteína es una proteína dCas9, y viceversa. El primer sitio del ácido nucleico diana se pone en proximidad de un segundo sitio de ácido nucleico diana debido a que las primera y segunda nucleasas proteicas catalíticamente inactivas están formando un complejo con el NATNA de tipo cruzado, es decir, son parte de la misma composición de ácido nucleico/proteína. En algunas realizaciones, la primera secuencia de ácido nucleico diana y la segunda secuencia de ácido nucleico diana están en polinucleótidos separados (por ejemplo, en diferentes cromosomas) o un único polinucleótido comprende la primera secuencia de ácido nucleico diana y la segunda secuencia de ácido nucleico diana (por ejemplo, en secciones diferentes del mismo cromosoma).

En otra realización más, la presente invención también incluye método de modulación de la transcripción *in vitro* o *in vivo*, por ejemplo, la transcripción de un gen que comprende secuencias de elementos reguladores, en el que el método no comprende un método de tratamiento del cuerpo humano o animal. El método comprende poner en contacto una primera secuencia de ácido nucleico diana y una segunda secuencia de ácido nucleico diana con una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas en un complejo con una primera proteína con la que el NATNA de tipo cruzado forma un

complejo y una segunda proteína con la que el NATNA de tipo cruzado forma un complejo (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, una proteína Cpfl con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo, y una proteína Cas9 con la que el NATNA de tipo cruzado con CRISPR de Clase 2 forma un complejo), facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia de ácido nucleico diana y la segunda secuencia de ácido nucleico diana. Al menos una de entre la primera secuencia de ADN diana y la segunda secuencia de ADN diana comprende las secuencias de elementos reguladores. La primera secuencia de ADN diana es complementaria de un primer espaciador NATNA (por ejemplo, un elemento espaciador Cpfl-NATNA o un elemento espaciador Cas9-NATNA), en el que la primera proteína asociada es un nucleasa proteica catalíticamente inactiva (por ejemplo, una proteína dCpfl o una proteína dCas9, respectivamente). La segunda secuencia de ADN diana es complementaria de un segundo espaciador NATNA (por ejemplo, un elemento espaciador Cpfl-NATNA o un elemento espaciador Cas9-NATNA), en el que la segunda proteína asociada es una nucleasa proteica catalíticamente inactiva (por ejemplo, una proteína dCpfl o una proteína dCas9, respectivamente). Por ejemplo, si la primera proteína es una proteína dCpfl, la segunda proteína es una proteína dCas9, y viceversa. Además, la primera y/o segunda proteína pueden ser proteínas de fusión, por ejemplo, una dCas9 fusionada a un dominio represor o activador, y/o dCpfl fusionada con un dominio represor o activador. La unión de la composición de ácido nucleico/proteína a la primera secuencia de ADN diana y la segunda secuencia de ADN diana modula la transcripción del gen. En algunas realizaciones, la primera secuencia de ADN diana y la segunda secuencia de ADN diana comprende las secuencias de elementos reguladores, y la primera secuencia de ADN diana comprende un promotor y la segunda secuencia de ADN diana comprende un sitio de inicio de la transcripción.

Cualquiera de los componentes de las composiciones de ácido nucleico/proteína comprende un NATNA de tipo cruzado modificado de la presente invención como se define en las reivindicaciones adjuntas o las secuencias de ácido nucleico que codifica dichos componentes, como se ha descrito anteriormente, se pueden incorporar en un kit, incluyendo opcionalmente uno o más reactivos. En algunas realizaciones, un kit incluye un paquete con uno o más recipientes que albergan los elementos del kit, como una o más composiciones por separado u, opcionalmente, como una mezcla en la que se permitirá la compatibilidad de los componentes. En algunas realizaciones, los kits también comprenden un tampón, un agente tampón, una sal, una solución acuosa estéril, y/o conservantes. Los kits ilustrativos comprenden un NATNA de tipo cruzado con CRISPR de Clase 2 modificado y opcionalmente una proteína Cpfl y/o una proteína Cas9, y una o más secuencias de ácido nucleico que codifican un NATNA de tipo cruzado con CRISPR de Clase 2 modificado, y opcionalmente una o más secuencias de ácido nucleico que codifiquen una proteína Cpfl y/o una proteína Cas9.

Además, los kits pueden comprender adicionalmente instrucciones para la utilización de los componentes de las composiciones de ácido nucleico/proteína que comprenden un NATNA de tipo cruzado modificado de la presente invención como se define en las reivindicaciones adjuntas o secuencias de ácido nucleico que codifica dichos componentes. Las instrucciones incluidas en los kits de la invención se pueden fijar en el material de empaquetado o se pueden incluir como un prospecto. Aunque las instrucciones son normalmente materiales escritos o impresos, no se limitan a estos. Se contempla cualquier medio capaz de almacenar dichas instrucciones y comunicarlas a un usuario final en la presente invención. Dichos medios incluyen, pero no se limitan a, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, un CD ROM), etiquetas RF y similares. Las instrucciones también pueden incluir la dirección de un sitio en internet que proporciona las instrucciones.

En otro aspecto, se desvelan métodos de producción o fabricación de un NATNA de tipo cruzado modificado o una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado de la presente invención. En un aspecto, los métodos de producción o fabricación comprenden la síntesis química de un NATNA de tipo cruzado modificado. En algunos aspectos el NATNA de tipo cruzado modificado comprende bases de ARN y se pueden generar a partir de armazones de ADN utilizando la transcripción *in vitro*.

Una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado) puede comprender adicionalmente un marcador detectable, que incluye un resto que puede proporcionar una señal detectable. Ejemplos de marcadores detectables incluyen, pero no se limitan a, una enzima, un radioisótopo, un miembro de una pareja de unión específica, un fluoróforo (FAM), una proteína fluorescente (proteína fluorescente verde, proteína fluorescente roja, mCherry, tdTomato), un aptámero de ADN o ARN junto con un fluoróforo adecuado (GFP aumentada (EGFP), "Espinaca"), un punto cuántico, un anticuerpo y similares. Un gran número y variedad de marcadores detectables adecuados son bien conocidos por el experto habituado en la técnica.

Una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado como se define en las reivindicaciones adjuntas (por ejemplo, un NATNA de tipo cruzado con CRISPR de Clase 2 modificado) o las células modificadas para su uso en una composición de ácido nucleico /proteína que comprende un NATNA de tipo cruzado modificado como se describe en el presente documento, se puede utilizar como una composición farmacéutica formulada, por ejemplo, con un excipiente farmacéuticamente aceptable. Los excipientes ilustrativos incluyen vehículos, estabilizantes, diluyentes, agentes de dispersión, agentes de suspensión, agentes espesantes, y similares. La composición farmacéutica puede facilitar la administración de una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado a un organismo. Las composiciones farmacéuticas se pueden

administrar en cantidades terapéuticamente eficaces mediante distintas formas y vías que incluyen, por ejemplo, la administración intravenosa, subcutánea, intramuscular, oral, en aerosol, parenteral, oftálmica y pulmonar.

5 Se obtienen numerosas ventajas utilizando un NATNA de tipo cruzado modificado de la presente invención como se define en las reivindicaciones adjuntas, que comprenden un primer NATNA y un segundo NATNA, en el que una primera proteína es capaz de formar un complejo con el primer NATNA del NATNA de tipo cruzado, y una segunda proteína es capaz de formar un complejo con el segundo NATNA del NATNA de tipo cruzado (por ejemplo un NATNA de tipo cruzado con CRISPR de Clase 2 modificado que comprende un Cpf1-NATNA y un Cas9-NATNA, en el que una proteína Cpf1 es capaz de formar un complejo con el Cpf1-NATNA del NATNA de tipo cruzado con CRISPR de Clase 2, y una proteína Cas9 es capaz de formar un complejo con el Cas9-NATNA del NATNA de tipo cruzado con CRISPR de Clase 2) que incluye pero no se limita a:

- 15 • una reducción en la unión de direccionamiento utilizando una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA) con respecto a un complejo de primer NATNA/primer proteína solo (por ejemplo, un complejo Cpf1-NATNA/proteína Cpf1) o un complejo segundo NATNA/segunda proteína solo (por ejemplo, un complejo Cas9-NATNA/proteína Cas9);
- 20 • la ligadura de un polinucleótido donante mediante el uso de una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA) para poner el polinucleótido donante en proximidad de un corte en un ácido nucleico de cadena doble;
- 25 • poner en proximidad dos polinucleótidos separados (por ejemplo, en dos cromosomas diferentes) o dos regiones de un único polinucleótido (por ejemplo, en dos regiones de un único cromosoma) entre ellos utilizando una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA); y
- la modulación de la transcripción de un gen diana uniendo una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA) a secuencias reguladoras unidas operativamente al gen diana.

Otra ventaja más de una composición de ácido nucleico/proteína comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA) es la eliminación de la necesidad de una secuencia PAM para la escisión de un ácido nucleico diana. Por ejemplo, se selecciona una primera secuencia de ácido nucleico diana complementaria de un primer elemento espaciador (véase, por ejemplo, la FIG. 13A, 1305). La primera secuencia de ácido nucleico diana comprende una secuencia PAM para la primera proteína (por ejemplo, un Cpf1 PAM). Se selecciona una segunda secuencia de ácido nucleico diana complementaria de un segundo elemento espaciador (véase, por ejemplo, la FIG. 13A, 1306), en el que la segunda secuencia de ácido nucleico diana carece de una secuencia PAM para la segunda proteína (por ejemplo, una Cas9 PAM) y está en estrecha proximidad a la primera secuencia de ácido nucleico de direccionamiento. La primera proteína (por ejemplo, una proteína Cpf1) en una composición de ácido nucleico/proteína que comprende un NATNA de tipo cruzado modificado (por ejemplo, un Cas9-Cpf1-NATNA) es de endonucleasa inactiva (por ejemplo, una proteína dCpf1) y la segunda proteína tiene una endonucleasa activa (por ejemplo, una proteína Cas9). El no enrollamiento local causado por la endonucleasa proteica inactiva que se unen al primer ácido nucleico diana puede permitir la escisión en el segundo sitio del ácido nucleico diana incluso en ausencia de un PAM para la endonucleasa proteica activa. (De manera similar se puede utilizar una endonucleasa inactiva Cas9 y una endonucleasa activa Cpf1).

45 Se contemplan distintas realizaciones en el presente documento que incluyen, pero no se limitan a una o más de las siguientes. Las realizaciones se numeran para la facilidad de referencia.

Realización 1. Un ácido nucleico de tres elementos modificado que se dirige a ácidos nucleicos (NATNA), que comprende: un NATNA asociado con Cpf1 que comprende un elemento espaciador, un extremo 5' y un extremo 3'; un primer NATNA asociado con Cas9 que comprende un elemento espaciador, un extremo 5', y un extremo 3'; y un segundo NATNA asociado con Cas9 que comprende un elemento tracr, un extremo 5' y un extremo 3'; en el que el primer NATNA asociado con Cas9 y el segundo NATNA asociado con Cas9 se asocian mediante enlace de hidrógeno de pares de bases; y en el que el NATNA asociado con Cpf1 está conectado con el extremo 5' o el extremo 3' del primer NATNA asociado con Cas9 que comprende un elemento espaciador, o el extremo 5' o el extremo 3' del segundo NATNA asociado con Cas9 que comprende un elemento tracr.

55 Realización 2. El NATNA de tres elementos modificado de la realización 1, en el que un complejo formado entre el NATNA asociado con Cpf1 y una proteína Cpf1 es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble; y en el que un complejo formado entre el primer NATNA asociado con Cas9, el segundo NATNA asociado con Cas9, y una proteína Cas9 es capaz de unirse a una segunda secuencia de ácido nucleico diana de cadena doble.

60 Realización 3. El NATNA de tres elementos modificado de la realización 1 o 2, en el que el NATNA asociado con Cpf1 se conecta con el extremo 5' o el extremo 3' del primer NATNA asociado con Cas9, o el extremo 5' o el extremo 3' del segundo NATNA asociado con Cas9, y la conexión es mediante enlaces de hidrógeno.

Realización 4. El NATNA de tres elementos modificado de la realización 3, en el que el NATNA asociado a Cpf1 comprende adicionalmente un elemento enlazador conectado covalentemente en el extremo 5' y/o el extremo 3' del NATNA asociado con Cpf1.

65 Realización 5. El NATNA de tres elementos modificado de la realización 3 o 4, en el que el primer NATNA asociado

con Cas9 y/o el segundo NATNA asociado con Cas9 que comprende adicionalmente un elemento enlazador conectado covalentemente al extremo 5' y/o el extremo 3' del primer NATNA asociado con Cas9 y/o el segundo NATNA asociado con Cas9.

5 Realización 6. El NATNA de tres elementos modificado de la realización 1 o 2, en el que el NATNA asociado con Cpfl se conecta con el extremo 5' o el extremo 3' del primer NATNA asociado con Cas9, o el extremo 5' o el extremo 3' del segundo NATNA asociado con Cas9, y la conexión es mediante enlaces de hidrógeno.

Realización 7. El NATNA de tres elementos modificado de la realización 6, en el que el NATNA asociado a Cpfl comprende adicionalmente un elemento enlazador conectado covalentemente al extremo 5' y/o el extremo 3' del segundo NATNA asociado con Cpfl.

10 Realización 8. El NATNA de tres elementos modificado de la realización 6 o 7, en el que el primer NATNA asociado con Cas9 y/o el segundo NATNA asociado con Cas9 comprende adicionalmente un elemento enlazador conectado covalentemente en el extremo 5' y/o el extremo 3' del primer NATNA asociado con Cas9 y/o el segundo NATNA asociado con Cas9.

15 Realización 9. El NATNA de tres elementos modificado de la realización 1 o 2, en el que el extremo 3' del primer NATNA asociado con Cas9 está conectado covalentemente mediante un elemento bucle con el extremo 5' del segundo NATNA asociado con Cas9 formando un NATNA asociado con Cas9 sencillo que tiene un extremo 5' y un extremo 3'; y en el que el NATNA asociado con Cpfl está conectado con el extremo 5' o el extremo 3' del NATNA asociado con Cas9 sencillo, y la conexión es mediante enlaces de hidrógeno.

20 Realización 10. El NATNA de tres elementos modificado de la realización 9, en el que el NATNA asociado con Cpfl comprende adicionalmente un elemento enlazador conectado covalentemente en el extremo 5' y/o el extremo 3' del NATNA asociado con Cpfl.

Realización 11. El NATNA de tres elementos modificado de la realización 9 o 10, en el que el NATNA asociado con Cas9 sencillo comprende adicionalmente un elemento enlazador conectado covalentemente al extremo 5' y/o el extremo 3' del NATNA asociado con Cas9 sencillo.

25 Realización 12. El NATNA de tres elementos modificado de la realización 1 o 2, en el que el extremo 3' del primer NATNA asociado con Cas9 está conectado covalentemente mediante un elemento bucle con el extremo 5' del segundo NATNA asociado con Cas9 formando un NATNA asociado con Cas9 sencillo; y en el que el NATNA asociado con Cpfl se conecta con el extremo 5' o el extremo 3' del ácido nucleico que se dirige a ácidos nucleicos asociado con Cas9 sencillo, y la conexión es mediante un enlace covalente.

30 Realización 13. El NATNA de tres elementos modificado de la realización 12, en el que el NATNA asociado con Cpfl comprende adicionalmente un elemento enlazador conectado covalentemente al extremo 5' y/o el extremo 3' del NATNA asociado con Cpfl.

Realización 14. El NATNA de tres elementos modificado de la realización 12 o 13, en el que el NATNA asociado con Cas9 sencillo comprende adicionalmente un elemento enlazador conectado al extremo 5' y/o el extremo 3' del NATNA asociado con CAS9 sencillo.

Realización 15. El NATNA de tres elementos modificado de cualquiera de las realizaciones anteriores, en el que al menos uno de los siguientes elementos comprende un ARN: un NATNA asociado con Cpfl; un primer NATNA asociado con Cas9; o un segundo NATNA asociado con Cas9.

40 Realización 16. El NATNA de tres elementos modificado de cualquiera de las realizaciones anteriores, en el que al menos uno de los siguientes elementos comprende un ADN: un NATNA asociado con Cpfl; un primer NATNA asociado con Cas9; o un segundo NATNA asociado con Cas9.

Realización 17. Una composición de ácido nucleico/proteína que comprende:

el NATNA de tres elementos modificado de cualquiera de las realizaciones anteriores, una proteína Cas9, y una proteína Cpfl.

45 Realización 18. La composición de ácido nucleico/proteína de la realización 17, en el que el NATNA de tres elementos modificado de la realización forma un complejo con la proteína Cas9 y la proteína Cpfl.

Realización 19. La composición de ácido nucleico/proteína de la realización 18, en la que la proteína Cpfl es inactiva para la actividad endonucleasa.

50 Realización 20. La composición de ácido nucleico/proteína de la realización 19, que comprende adicionalmente un polinucleótido donante en el que el polinucleótido donante comprende una secuencia de nucleótidos complementaria del elemento espaciador de NATNA asociado con Cpfl.

Realización 21. La composición de ácido nucleico/proteína de la realización 20, en la que el elemento espaciador de NATNA asociado con Cpfl se conecta con el polinucleótido donante mediante enlaces de hidrógeno entre la secuencia de nucleótidos del polinucleótido donante complementaria del elemento espaciador y el elemento espaciador.

55 Realización 22. La composición de ácido nucleico/proteína de la realización 18, en el que la proteína Cas9 es inactiva para las actividades de endonucleasa.

Realización 23. La composición de ácido nucleico/proteína de la realización 22, que comprende adicionalmente un polinucleótido donante en el que el polinucleótido donante comprende una secuencia complementaria al elemento espaciador del primer NATNA asociado con Cas9.

60 Realización 24. La composición de ácido nucleico/proteína de la realización 23, en el que el primer elemento espaciador de NATNA asociado con Cas9 se conecta con el polinucleótido donante por enlaces de hidrógeno entre la secuencia de nucleótidos del polinucleótido donante complementara del elemento espaciador y el elemento espaciador.

65 Realización 25. La composición de ácido nucleico/proteína de la realización 18, en el que la proteína Cas9 es inactiva para las actividades endonucleasa y la proteína Cpfl es inactiva para la actividad endonucleasa.

Realización 26. Uno o más vectores comprenden: secuencias de nucleótidos que codifican el NATNA de tres elementos modificado de cualquiera de las realizaciones 1 a 16.

Realización 27. Una célula recombinante que comprende: el uno o más vectores de la realización 26.

5 Realización 28. Una célula recombinante que comprende: el NATNA de tres elementos modificado de la realización de cualquiera de las realizaciones 1 a 16.

Realización 29. Una célula recombinante que comprende: La composición de ácido nucleico/proteína de una cualquiera de las realizaciones 17 a 25.

10 Realización 30. La célula recombinante de cualquiera de las realizaciones 17 a 29, en el que la célula es una célula vegetal, una célula de levadura, una célula bacteriana, una célula de insecto, una célula de alga, o una célula de mamífero.

Realización 31. Un método de modificación de ADN que comprende:

15 poner en contacto una primera secuencia de ADN diana y una segunda secuencia de ADN diana del ADN con la composición de ácido nucleico/proteína de la realización 18, en el que la primera secuencia de ADN diana es complementaria del primer elemento espaciador del NATNA asociado con Cas9, la segunda secuencia de ADN diana es complementaria del elemento espaciador del NATNA asociado con Cpf1, la composición de ácido nucleico/proteína se une a la primera secuencia de ADN diana y la segunda secuencia de ADN diana, y la proteína Cas9 y la proteína Cpf1 cortan la primera secuencia de ADN diana y la segunda secuencia de ADN diana, respectivamente, dando como resultado una modificación del ADN, en el que el método no comprende un proceso para la modificación de la identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal.

20

25 Realización 32. Un método de modificación del ADN que comprende:

poner en contacto una primera secuencia de ADN diana de un polinucleótido donante y una secuencia de ADN diana de un polinucleótido donante con la composición ácido nucleico/proteína de la reivindicación 21, en el que la primera secuencia de ADN diana es complementaria del primer elemento espaciador del primer NATNA asociado con Cas9, la segunda secuencia de ADN diana es complementaria del elemento espaciador del NATNA asociado con Cpf1, la composición de ácido nucleico/proteína se une a la primera secuencia de ADN diana, la proteína Cas9 corta la primera secuencia de ADN diana, y al menos una parte del polinucleótido donante se introduce en el ADN dando como resultado una modificación del ADN, en el que el método no comprende un proceso para la modificación de identidad genética de la línea germinal de un ser humano y, en el que en el que el método no comprende un método de tratamiento del cuerpo humano o animal.

30

35

Realización 33. Un método de modificación del ADN que comprende:

40 poner en contacto una primera secuencia de ADN diana de un polinucleótido donante y una segunda secuencia de ADN diana con la composición de ácido nucleico/proteína de la realización 24, en el que la primera secuencia de ADN diana es complementaria del elemento espaciador del primer NATNA asociado con Cas9, la segunda secuencia de ADN diana es complementaria del elemento espaciador del NATNA asociado con Cpf1, la composición de ácido nucleico/proteína se une a la segunda secuencia de ADN diana, la proteína Cpf1 corta la segunda secuencia de ADN diana, y al menos una parte del polinucleótido donante se introduce en el ADN dando como resultado una modificación del ADN, en el que el método no comprende un proceso para la modificación de identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal.

45

50 Realización 34. Un método de poner un primer sitio de ADN diana en proximidad de un sitio de ADN diana que comprende:

poner en contacto una primera secuencia de ADN diana y una segunda secuencia de ADN diana con la composición de ácido nucleico/proteína de la realización 25, en el que la primera secuencia de ADN diana es complementaria del elemento espaciador del primer NATNA asociado con Cas9, la segunda secuencia de ADN diana es complementaria del elemento espaciador del NATNA asociado con Cpf1, la composición de ácido nucleico/proteína se une a la primera secuencia de ADN diana y la segunda secuencia de ADN diana poniendo el primer sitio de ADN diana en proximidad de la segunda secuencia de ADN diana, en el que el método no comprende un proceso para la modificación de identidad genética de la línea germinal de un ser humano y, en el que el método no comprende un método de tratamiento del cuerpo humano o animal.

Realización 35. El método de la realización 34, en el que la primera secuencia de ADN diana y la segunda secuencia de ADN diana están en polinucleótidos separados.

Realización 36. Un método de modulación de la transcripción de un gen que comprende secuencias de elemento regulador, que comprende:

poner en contacto una primera secuencia de ADN diana y una segunda secuencia de ADN diana con la composición de ácido nucleico/proteína de la realización 25, en el que la primera secuencia de ADN diana es complementaria del elemento espaciador del primer NATNA asociado con Cas9, la segunda secuencia de ADN diana es complementaria del elemento espaciador del NATNA asociado con Cpf1, al menos una de entre la primera secuencia de ADN diana o la segunda secuencia de ADN diana comprende las secuencias de elemento regulador, la composición de ácido nucleico/proteína se une a la primera secuencia de ADN diana y la segunda secuencia de ADN diana, y la unión de la composición de ácido nucleico/proteína modula la transcripción del gen, en el que el método no comprende un método de tratamiento del cuerpo humano o animal.

## Experimentación

Aspectos de la presente invención se ilustran en los siguientes Ejemplos. Se ha hecho un esfuerzo para asegurar la precisión con respecto a los números utilizados (por ejemplo, cantidades, concentraciones, porcentaje de cambio, y similares) pero se debería contar con algunos errores experimentales y desviaciones. A menos de que se indique otra cosa, la temperatura es en grados centígrados y la presión es o está cerca de la atmosférica. Se debería entender que estos Ejemplos se dan solamente a modo de ilustración y no pretenden limitar el alcance que los inventores consideran como distintos aspectos de la presente invención.

## Materiales y métodos

Las secuencias de secuencias de oligonucleótido se proporcionaron por fabricantes comerciales para la síntesis.

Los ácidos nucleicos de tipo cruzado modificados que se dirigen a ácidos nucleicos ("NATNA de tipo cruzado"; por ejemplo, ácidos nucleicos Cas9-Cpf1 modificados que se dirigen a ácidos nucleicos ("Cas9-Cpf1-NATNA")) se ensamblaron por PCR utilizando cebadores de solapamiento 3' que contenían secuencias de ADN que se corresponde con un NATNA de tipo cruzado.

### Ejemplo 1

#### Producción de componentes del ácido nucleico de tipo cruzado que se dirige a ácidos nucleicos

Este ejemplo describe la producción de NATNA de tipo cruzado, por ejemplo, un Cas9-Cpf1-NATNA similar al Cas9-Cpf1-NATNA que se ilustra en la FIG. 4. El Cas9-Cpf1-NATNA se diseñó con un espaciador que se dirige a las secuencias codificantes del factor A de crecimiento vascular endotelial (VEGFA) incorporado corriente arriba (por ejemplo, en una dirección 5') al componente de estructura del Cas9-sgARN y un espaciador que se dirige a las secuencias codificantes de una indolamina 2,3-dioxigenasa 2 (IDO2) incorporado corriente abajo (por ejemplo en una dirección 3') al componente de estructura del Cpf1 ARN guía. El Cas9-Cpf1-NATNA se diseñó de manera que el extremo 3' de la secuencia del Cas9 ARN guía se unía (mediante un enlace covalente) al extremo 5' de la secuencia Cpf1 ARN guía para crear un único polinucleótido con una estructura de polinucleótido continua (por ejemplo, FIG. 4).

Los componentes de ARN se produjeron por transcripción *in vitro* (por ejemplo, con el kit de síntesis de ARN de alto rendimiento T7 Quick, New England Biolabs, Ipswich, MA) a partir de una matriz de ADN de cadena doble que incorpora un promotor T7 en el extremo 5' de las secuencias de ADN.

La matriz de ADN de cadena doble para el componente Cas9-Cpf1-NATNA específico ("Cas9-Cpf1-NATNA<sub>1</sub>") utilizado en los siguientes ejemplos se ensambló por PCR utilizando cebadores de solapamiento 3' que contenían las secuencias de ADN correspondientes al componente Cas9-Cpf1-NATNA<sub>1</sub>. Los oligonucleótidos utilizados en el ensamblaje se presentan en la Tabla 1.

Tabla 1

Cebadores de solapamiento para la generación de Cas9-Cpfl-NATNA <sub>1</sub> -Matriz codificante	
SEQ ID NO.	Secuencia*
SEQ ID NO. 1	AGTGTAAATACGACTCACTATAG
SEQ ID NO. 2	TAATACGACTCACTATAGGGGTGGGGGGAGTTTGCTCCGT TTTAGAGCTAGAAATAGCAAAGTTGAGATAAGGCTAGTCCG TTATCAACTTG
SEQ ID NO. 3	CCAGGCACTGCTTCTTTCTCTACCATCTACAACAGTAGAA ATTAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAAC GGACTAGCCTT
SEQ ID NO. 4	CCAGGCACTGCTTCTTTCTCTACC
* los espaciadores están subrayados	

Los cebadores de ADN estaban presentes a una concentración de 2 nM cada uno. Los dos cebadores de ADN externos se corresponden al promotor T7 (SEQ ID NO. 1) y el extremo 3' de la secuencia de ARN (SEQ ID NO. 4) se utilizaron a 640 nM para dirigir la reacción de amplificación. Las reacciones PCR se llevaron a cabo utilizando la mezcla Q5 Hot Start High-Fidelity 2X Master (New England Biolabs, Ipswich, MA) siguiendo las instrucciones del fabricante. Las reacciones de ensamblaje por PCR se llevaron a cabo utilizando las siguientes condiciones de ciclo térmico: 98 °C durante 2 minutos, 35 ciclos de 15 segundos a 98 °C, 15 segundos a 60 °C, 15 segundos a 72 °C, y una extensión final a 72 °C durante 2 minutos. Se evaluó la calidad del ADN mediante electroforesis en gel de agarosa (1,5 % SYBR® Safe; Life Technologies, Grand Island, NY).

Se transcribieron entre 0,26-0,5 µg de la matriz de ADN para el componente Cas9-Cpfl-NATNA<sub>1</sub> utilizando el kit de síntesis de ARN de alto rendimiento T7 (New England Biolabs, Ipswich, MA) durante aproximadamente 16 horas a 37 °C. Las reacciones de transcripción se trataron con DNasa I (New England Biolabs, Ipswich, MA) y se purificaron utilizando el kit de concentración y limpieza de ARN GeneJet (Life Technologies, Grand Island, NY). El rendimiento de ARN se cuantificó utilizando un Sistema Nanodrop™ 2000 (Thermo Scientific, Wilmington, DE). La calidad del ARN transcrito se comprobó por electroforesis en gel de agarosa (un 2 % de SYBR® Safe; Life Technologies, Grand Island, NY). La secuencia de Cas9-Cpfl-NATNA<sub>1</sub> se muestra en la Tabla 2.

Tabla 2

Secuencia de Cas9-Cpfl-NATNA <sub>1</sub>	
SEQ ID NO.	Secuencia *
SEQ ID NO. 5	GGGUGGGGGGAGUUUGC <u>UCCGUUUU</u> AGAGCUAGAAUAG CAAGUUGAGAU <u>AAGGCUAGUCCGUUAUCA</u> CUUGAAAAA GUGGCACCGAGUCGGUGCUAAUUUCUACUGUUGUAGAU <u>GGUAGAGAAAGAAGCAGUGCCUGG</u>
* los espaciadores están subrayados	

Este método para la producción de Cas9-Cpfl-NATNA<sub>1</sub> se puede aplicar a la producción de otros NATNA de tipo cruzado descritos en el presente documento.

## Ejemplo 2

### Producción de secuencias de ADN diana de cadena doble para su uso en los ensayos de escisión

Se produjeron las secuencias de ADN diana de cadena doble para su uso en los ensayos de escisión Cas *in vitro* utilizando amplificación por PCR de las secuencias de ácido nucleico diana seleccionadas del ADN genómico humano.

Se amplificaron las secuencias de ADN diana (por ejemplo, VEGFA e IDO2) para los ensayos bioquímicos por PCR a partir del ADN genómico (gADN) de la línea celular K562 humana (Colección Americana de Cultivos Tipo (ATCC), Manassas, VA) preparada con fenol-cloroformo. Las reacciones se llevaron a cabo con la Mezcla Q5 Hot Start High-Fidelity 2X Master (New England Biolabs, Ipswich, MA) siguiendo las instrucciones del fabricante. Se utilizaron 20 ng/µl de gADN con un volumen final de 25 µl para amplificar la secuencia de ácido nucleico diana seleccionada en las siguientes condiciones: 98 °C durante 2 minutos, 35 ciclos de 15 segundos a 98 °C, 15 segundos a 60 °C, 15 segundos a 72 °C, y una extensión final a 72 °C durante 2 minutos. Los productos de la PCR se purificaron utilizando los tubos de purificación Spin Smart™ PCR (Denville Scientific, South Plainfield, NJ) y se cuantificaron utilizando un

espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE).

Los cebadores directos e inversos utilizados para la amplificación de las secuencias dirigidas seleccionadas del gADN se presentan en la Tabla 3.

5

Tabla 3

Secuencias de cebador de la secuencia de ADN diana de cadena doble	
SEQ ID NO.	Secuencia
SEQ ID NO. 6	CACTCTTCCCTACACGACGCTCTCCGATCTCCAGATGGC ACATTGTCAGA
SEQ ID NO. 7	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTAGTACTGCC GTCTGC
SEQ ID NO. 8	CACTCTTCCCTACACGACGCTCTCCGATCTAAGGAACCT GGAGACCATCA
SEQ ID NO. 9	GGAGTTCAGACGTGTGCTCTTCCGATCTGAAAGGCACTGA GTGGGAAG

Las secuencias diana de VEGFA se amplificaron utilizando la SEQ ID NO. 6 y la SEQ ID NO. 7, dando lugar a una secuencia de ADN diana de doble cadena de 261 pb. Las secuencias de ADN de IDO2 se amplificaron utilizando la SEQ ID NO. 8, y la SEQ ID NO. 9, dando lugar a una secuencia de ADN diana de cadena doble de 232 pb.

10

se pueden obtener otras secuencias de ADN diana de cadena doble utilizando esencialmente el mismo método. Para las secuencias de ácido nucleico diana no humanas, se puede utilizar el ADN genómico del organismo seleccionado (por ejemplo, plantas, bacterias, levaduras, algas) en vez del ADN derivado de células humanas. Además, se pueden utilizar fuentes de polinucleótido distintas del ADN genómico (por ejemplo, vectores y fragmentos de ADN aislados en gel).

15

### Ejemplo 3

#### 20 Ensayos de escisión con Cas

Este Ejemplo ilustra el uso de los NATNA de tipo cruzado/proteína Cas en los ensayos de escisión.

En este ejemplo, se utilizaron un complejo Cas9-Cpfl-NATNA1/proteína Cas9 y una complejo Cas9-Cpfl-NATNA1/proteína Cpfl en ensayos de escisión con Cas9 o Cpfl *in vitro*, respectivamente, para evaluar y comparar el porcentaje de escisión de los complejos Cas9-Cpfl-NATNA1/proteína Cas9 y complejos Cas9-Cpfl-NATNA1/proteína Cpfl con respecto a las secuencias de ADN diana de cadena doble.

25

La escisión de secuencias de ADN diana de cadena doble se determinó utilizando los componentes del Cas9-Cpfl-NATNA1 del Ejemplo 1, con una o más proteínas Cas equivalentes para formar un complejo, contra una secuencia de ADN diana de cadena doble (VEGFA e IDO2; Ejemplo 2).

30

Se expresaron de manera recombinante la Cas9 de *S. pyogenes* y Cpfl de *F. novicida* en *E. coli* y se purificaron para su uso en el ensayo de escisión bioquímica *in vitro*.

35

Los componentes de Cas9-Cpfl-NATNA1 se diluyeron hasta una concentración de trabajo adecuada y se incubaron durante 2 minutos a 95 °C, se retiraron del termociclador y se permitió que se equilibraran a temperatura ambiente.

Se añadió el Cas9-Cpfl-NATNA1 a cualquiera de una mezcla de reacción Cas9 o una mezcla de reacción Cpfl. La mezcla de reacción Cas9 comprendía la proteína Cas diluida hasta una concentración final de 20 nM en tampón de reacción (20 nM de HEPES, 100 mM de KCl, 5 mM de MgCl<sub>2</sub>, 1 mM de DTT, y un 5 % de glicerol a un pH de 7,4). La mezcla de reacción Cpfl comprende la proteína Cpfl diluida hasta una concentración final de 50 nM en tampón de reacción (20 nM de HEPES, 100 mM de KCl, 5 mM de MgCl<sub>2</sub>, 1 mM de DTT, y un 5 % de glicerol a un pH de 7,4). En la mezcla de reacción, la concentración final de Cas9-Cpfl-NATNA1 era la siguientes: 200 nM en la mezcla de reacción Cas9 y 500 nM en la mezcla de reacción Cpfl. Cada mezcla de reacción se incubó a 37 °C durante 10 minutos. La reacción de escisión se inició por adición de la secuencia de ADN diana hasta una concentración final de 10 nM. Las muestras se mezclaron y centrifugaron brevemente antes de incubarse durante 15 minutos a 37 °C. Las reacciones de escisión se terminaron por adición de Proteinasa K (Denville Scientific, South Plainfield, NJ) a una concentración final de 0,2 µg/µl y 0,44 mg/µl de solución de RNasa A (Sigma Aldrich, St. Louis, MO). Las muestras se incubaron entonces durante 25 minutos a 37 °C y 25 minutos a 55 °C. Se evaluaron 12 µl de la reacción total en cuanto a la actividad de escisión por electroforesis en gel de agarosa (un 2 % de SYBR® Gold; Life Technologies, Grand Island, NY). Para la escisión con Cas9 de una secuencia de ADN diana de doble cadena del VEGFA, la aparición de bandas

40

45

50

de ADN aproximadamente a 158 pb y aproximadamente 103 pb indicaba que se había producido la escisión de la secuencia de ADN diana. Para la escisión con Cpf1 de una secuencia de ADN diana de cadena doble de IDO2, la aparición de bandas de ADN a aproximadamente 120 pb y aproximadamente 112 pb indicaba que se había producido la escisión de la secuencia de ADN diana. Se calcularon los porcentajes de escisión utilizando los valores del área bajo la curva (AUC) según se calculó por FIJI (ImageJ; un programa de procesamiento de imágenes de Java abierto) para cada fragmento de escisión y la secuencia de ADN diana, y dividiendo la suma de los fragmentos de escisión por la suma de los fragmentos de escisión y las secuencias de ADN diana.

5

10

La Tabla 4 presenta los resultados de los ensayos de escisión con Cas9 y Cpf1 utilizando secuencias de ácido nucleico diana de VEGFA y secuencias de ADN diana de cadena doble de IDO2.

Tabla 4

Escisión bioquímica de la secuencia de ADN diana con un Cas9-Cpf1-NATNA		
	Escisión de la secuencia de ADN diana de VEGFA con Cas9	Escisión de la secuencia de ADN diana de IDO2 con Cpf1
Cas9-Cpf1-NATNA <sub>1</sub>	93 %	68 %

15

Los datos presentados en la Tabla 4 demuestran que el Cas9-Cpf1-NATNA<sub>1</sub> facilitaba la escisión específica del sitio mediada por la proteína Cas de las secuencias de ADN diana de cadena doble. Los datos también demostraban que un Cas9-Cpf1-NATNA sencillo se puede utilizar con nucleasa Cas9 o nucleasa Cpf1 para facilitar la escisión dirigida de una secuencia de ADN diana de cadena doble.

20

Siguiendo las directrices de la presente memoria descriptiva, el ensayo de escisión bioquímica descrito en este ejemplo puede practicarlo un experto en la técnica con otros NATNA de tipo cruzado y sus proteínas Cas equivalentes (por ejemplo, proteínas Cas con CRISPR Tipo II o proteínas Cpf1 con CRISPR Tipo V).

#### Ejemplo 4

25

#### Análisis de secuenciación profunda para la detección de modificaciones de secuencia diana en células eucariotas

30

Este ejemplo ilustra el uso del análisis de secuenciación profunda para evaluar y comparar el porcentaje de escisión en las células de complejos NATNA de tipo cruzado/proteína Cas con respecto a las secuencias de ADN diana de cadena doble seleccionadas.

#### A. Diseño de un Cas9-Cpf1-NATNA

35

Se exploró el locus del factor 2 de crecimiento del receptor epidérmico humano (ERBB2) en cuanto a secuencias protoespaciadoras de Cas de 20 nucleótidos que se encuentran corriente arriba de una secuencia 5'-NGG PAM y una secuencia protoespaciadora de Cpf1 de 24 nucleótidos que se encuentra corriente abajo de una secuencia 3'-TTN PAM. Ejemplos de las secuencias protoespaciadoras seleccionadas por este método se muestran en la Tabla 5.

Tabla 5

Secuencias protoespaciadoras Cas9 y Cpf1 de ERBB2			
Nombre	SEQ ID NO.	Secuencia*	Coordenadas cromosómicas Hg38
Protoespaciador Cas9 ERBB2	SEQ ID NO. 10	GTGCCCTCGGTCACACTGTGIGG	chr17:39727008-39727030
Protoespaciador Cpf1 ERBB2	SEQ ID NO. 11	TTTCTCTGATGTTCCCTCAACT GTCACC	chr17:39727072-39727099

\* Los PAM específicos de nucleasa están subrayados

40

La secuencia espaciadora de Cas9 (es decir la secuencia protoespaciadora sin el PAM) se incorporó corriente arriba 5' de un componente armazón de Cas sgARN y la secuencia espaciadora de Cpf1 (es decir, la secuencia protoespaciadora sin el PAM) se incorporó corriente abajo 3' de un componente de armazón de Cpf1 ARN guía para formar el NATNA de tipo cruzado ERBB2-Cas9-Cpf1-NATNA. La secuencia de ARN para este ERBB2-Cas9-Cpf1-

NATNA se muestra en la Tabla 6.

Tabla 6

Secuencia ERBB2-Cas9-Cpfl-NATNA	
SEQ ID NO.	Secuencia*
SEQ ID NO. 12	<u>GUGCCCUCGGUCACACUGUGGUUUUAGAGCUAGAAAUAGCAA</u> <u>GUUGAGAUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA</u> <u>CCGAGUCGGUGCUAAAUUCUACUGUUGUAGAUUCUGAUGUU</u> <u>CCCUCAACUGUCACC</u>
* Los espaciadores de nucleasa están subrayados	

- 5 Siguiendo las directrices de la presente memoria descriptiva y Ejemplos, el experto habituado en la técnica puede diseñar NATNA de tipo cruzado adicionales.

#### B. Formación de Complejos nucleoproteicos ERBB2-Cas9-Cpfl-NATNA/proteína Cas9&Cpfl

- 10 Se encargaron los cebadores para ensamblar el ERBB2-Cas9-Cpfl-NATNA en un fabricante comercial, y se puede ensamblar el ERBB2-Cas9-Cpfl-NATNA esencialmente como se describe en el Ejemplo 1. La secuencia de ARN para un ERBB2-Cas9-Cpfl-NATNA se muestra en la Tabla 6. La Cas9 de *S. pyogenes* y Cpfl de *Acidaminococcus sp.* se marcaron ambos en el extremo C con dos secuencias de localización nuclear (NLS) y se expresaron recombinantemente en *E. coli*. Se formaron complejos ribonucleoproteicos (RNP) a una concentración de 40 pmol de proteína Cas9: 40 pmol de proteína Cpfl: 120 pmoles de ERBB2-Cas9-Cpfl-NATNA. Antes del ensamblaje con las proteínas Cas9 y Cpfl, el ERBB2-Cas9-Cpfl-NATNA se diluyó hasta la concentración deseada (120 pmol) a un volumen final de 2 µl se incubaron durante 2 minutos a 95 °C, se retiraron del termociclador, y se dejaron equilibrar a temperatura ambiente. Las proteínas Cas9 y Cpfl se diluyeron hasta una concentración en tampón de unión (20 mM de HEPES, 100 mM de KCl, 5 mM de MgCl<sub>2</sub>, 1 mM de DTT, y un 5 % de glicerol a un pH de 7,4) hasta un volumen final de 3 µl y se mezclaron con 2 µl de ERBB2-Cas9-Cpfl-NATNA seguido por una incubación a 37 °C durante 30 minutos.

#### C. Transfecciones celulares utilizando el RNP ERBB2-Cas9-Cpfl-NATNA/Cas9&Cpfl

- 25 Los complejos RNP se transfectaron en células HEK293 (ATCC Mannassas, VA), utilizando el Sistema Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ) y el siguiente protocolo. Los complejos RNP se dispensaron en un volumen final de 5 µl en pocillos individuales de una placa de 96 pocillos. El medio de cultivo celular se retiró de la placa de cultivo de células HEK293 y las células se despegaron con TrypLE™ (Thermo Fisher Scientific, Waltham, MA). Las células HEK293 suspendidas se aglomeraron por centrifugación durante 3 minutos a 200 x g, se aspiraron los reactivos TrypLE, y se lavaron las células con solución salina tampón de fosfato libre de magnesio y calcio (PBS). Las células se aglomeraron por centrifugación durante 3 minutos a 200 x g, se aspiró el PBS, y el aglomerado celular se re-suspende en 10 ml de PBS libre de magnesio y calcio.

- 35 Las células se pueden contar utilizando el Contador celular automático Countess® II (Life Technologies; Grand Island, NY). Se transfirieron  $2,2 \times 10^7$  células a un tubo de microfuga de 1,5 ml y se aglomeraron. El PBS se aspiró y las células se re-suspendieron en solución Nucleofector™ SF (Lonza, Allendale, NJ) hasta una densidad de  $1 \times 10^7$  células/ml. Se añaden entonces 20 µl de suspensión celular a cada pocillo individual que contenía 5 µl de complejos RNP, y todo el volumen de cada pocillo se transfirió a un pocillo de una Placa Nucleocuvette™ de 96 pocillos (Lonza, Allendale, NJ). La placa se cargó en el Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ) y las células se nucleofectaron utilizando el programa 96-CM-130 Nucleofector™ (Lonza, Allendale, NJ). Después de la nucleofección se añadieron 70 µl de medio de Eagle modificado de Dulbecco (DMEM: Thermo Fisher Scientific, Wilmington, DE) suplementado con un 10 % de Suero Fetal bovino (FBS: Thermo Fisher Scientific, Wilmington, DE), penicilina y estreptomycin (Life Technologies; Grand Island, NY) a cada pocillo y entonces se transfirieron 50 µl de la suspensión celular a una placa de cultivo celular de 96 pocillos que contenía 150 µl de medio de cultivo DMEM completo precalentado. La placa se transfirió entonces a una incubadora de cultivo tisular y se mantuvo a 37 °C en un 5 % de CO<sub>2</sub> durante 48 horas.

#### D. Generación de secuencia de ADN diana de cadena doble para la secuenciación profunda

- 50 Se aisló el gADN de las células HEK293 48 horas después de la transfección utilizando complejos RNP y se añadieron 50 µl de solución de extracción de ADN QuickExtract (Epicentre, Madison, WI) por pocillo, seguido por incubación a 37 °C durante 10 minutos, 65 °C durante 6 minutos y 95 °C durante 3 minutos para parar la reacción. El gADN aislado se diluyó entonces con 50 µl de agua estéril y se almacenaron las muestras a -80 °C.

- 55 Utilizando el gADN aislado, se llevó a cabo una primera PCR utilizando la mezcla Q5 Hot Start High-Fidelity 2X Master (New England Biolabs, Ipswich, MA) a una concentración de 1x, cebadores a 0,5 µM cada uno (SEQ ID NO. 13 y SEQ

ID NO. 14), 3,75 µl de gADN en un volumen final de 10 µl y se amplificaron a 98 °C durante 1 minuto, 35 ciclos de 10 s a 98 °C, 20 segundos a 60 °C, 30 segundos a 72 °C, y una extensión final a 72 °C durante 2 minutos. Los cebadores se diseñaron para amplificar la región del genoma dirigido por ERBB2-Cas9-Cpfl-NATNA. La reacción PCR se diluyó 1:100 en agua.

5 Se llevó a cabo la codificación de barras de la PCR utilizando una mezcla de reacción que comprendía la mezcla Q5 Hot Start High-Fidelity 2X Master (New England Biolabs, Ipswich, MA) a una concentración de 1x; cebadores a 0,5 µM cada uno (SEQ ID NO. 15 y SEQ ID NO. 15); 1 µl de una primera PCR diluida 1:100; en un volumen final de 10 µl; y la mezcla de reacción se amplificó a 98 °C durante 1 minuto, seguido por 13 ciclos de 10 s a 98 °C, 20 segundos a 60 °C, y 30 segundos a 72 °C, con una reacción de extensión final llevada a cabo a 72 °C durante 2 minutos.

#### E. Limpieza con SPRIselect

15 La reacción PCR se transfirió a un único tubo de microfuga para la limpieza de amplicones basado en perlas SPRIselect (Beckman Coulter, Pasadena, CA) para la secuenciación.

20 Para el amplicón, se añadieron 0,9x volúmenes de perlas SPRIselect, se mezclaron, y se incubaron a temperatura ambiente (TA) durante 10 minutos. El tubo microfuga se colocó en una plataforma de tubos magnética (Beckman Coulter, Pasadena, CA) hasta que la solución se aclaraba. Se retiró el sobrenadante y se desechó, y las perlas restantes se lavaron con 1 volumen de un 85 % de etanol, y se incubaron a TA durante 30 segundos. Después de la incubación, se aspiró el etanol y las perlas se secaron al aire a TA durante 10 minutos. el tubo de microfuga se retiró entonces de la plataforma magnética y se añadieron a las perlas 0,25x volúmenes de tampón Qiagen EB (Qiagen, Venlo, Holanda), se mezcló vigorosamente, y se incubó durante 2 minutos a temperatura ambiente. El tubo de microfuga se devolvió al imán, se incubó hasta que la solución se aclaró, y el sobrenadante que contenía los amplicones purificados se dispensó en un tubo de microfuga limpio El amplicón purificado se cuantificó utilizando un sistema Nanodrop 2000 (Thermo Scientific, Wilmington DE) y se analizó la calidad de la biblioteca utilizando un sistema Fragment Analyzer™ (Advanced Analytical Technologies, Ames, IA) y el Kit DNF-910 Double-stranded DNAREagent (Advanced Analytical Technologies, Ames, IA).

#### 30 F. Configuración de secuenciación profunda

El amplicón se normaliza a una concentración de 4 nM según se calcula por los valores Nanodrop y el tamaño de los amplicones. La biblioteca se analizó en un secuenciador MiSeq (Illumina, San Diego, CA) con el Kit MiSeq Reagent v2 (Illumina, San Diego, CA) durante 300 ciclos con dos ejecuciones de extremo emparejado de 151 ciclos más dos lecturas de índice de ocho ciclos.

#### G. Análisis de los datos de la secuenciación profunda

40 La identidad de productos en los datos de secuenciación se determinó basándose en las secuencias del código de barras adaptado en los amplicones en la ronda de codificación de barras de la PCR. Se utilizó una secuencia de comandos computacional para procesar los datos MiSeq que ejecuta, por ejemplo, las siguientes tareas:

- Las lecturas se alinean con el genoma humano (construcción GRCh38/38) utilizando el software Bowtie (bowtie-bio.sourceforge.net/index.shtml)
- 45 • Las lecturas alineadas se comparan con la secuencia del locus ERBB2 de tipo silvestre esperada, y las lecturas no alineadas a cualquier parte del locus ERBB2 se desechan.
- Las lecturas que coinciden con la secuencia ERBB2 de tipo silvestre se computan.
- Las lecturas con indels (inserciones o eliminaciones de bases) se categorizan por tipo de indel y se computan.
- 50 • Las lecturas indel totales se dividen por la suma de lecturas de tipo silvestre y las lecturas indel para dar lecturas del porcentaje de mutantes.

Mediante la identificación de las secuencias indel y las regiones marcadas por los complejos RNP de ERBB2-Cas9-Cpfl-NATNA/proteínas Cas9&Cpfl, se puede determinar el direccionamiento específico de secuencia en una línea celular humana. La distancia relativa entre los dos sitios de secuencia de ácido nucleico diana en el locus ERBB2 puede dar como resultado la escisión de la secuencia entre las secuencias de ácido nucleico diana de Cas9 y Cpfl.

55 Siguiendo las directrices de la presente memoria descriptiva y los Ejemplos, la edición en la célula de una secuencia genómica puede practicarse por un experto habituado en la técnica con otras proteínas Cas y sus NATNA de tipo cruzado equivalentes.

#### 60 **Ejemplo 5**

##### Identificación y exploración de crARN

Este Ejemplo describe un método para identificar crARN de Clase 2 en diferentes especies bacterianas.

65

Este Ejemplo describe un método por el que se identifican los crARN de especies que tienen un sistema CRISPR de Clase 2. El método que se presenta aquí se ha adaptado de Chylinski, K., et al, RNA Biology 10(5):726-37 (2013). No son necesarias las siguientes etapas para la exploración ni se debe seguir el orden de las etapas como se presentan.

5 A. Identificación de una especie que contiene un locus CRISPR de Clase 2

Utilizando una Herramienta de búsqueda de alineamiento local básico (BLAST, [blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)), se lleva a cabo una búsqueda de los genomas de distintas especies para identificar nucleasas Cas CRISPR de Clase 2 (por ejemplo, proteína Cas9, proteína Cpf1, proteínas tipo Cas9, proteínas tipo Cpf1, etc.). Los sistemas CRISPR de Clase 2 presentan una alta diversidad de secuencia a través de las especies, sin embargo, los ortólogos de nucleasa CRISPR de Clase 2 tienen dominios conservados, por ejemplo, un dominio endonucleasa HNH y/o un dominio RuvC/RNasa H. Los resultados del BLAST primarios se filtran para los dominios identificados, secuencias incompletas o truncadas se descartan y se identifican las especies que tienen ortólogos de nucleasa CRISPR de Clase 2.

15 Si se identifica un ortólogo de nucleasa CRISPR de Clase 2 en una especie, las secuencias adyacentes a la secuencia codificante del ortólogo de la proteína Cas (por ejemplo, una proteína Cas9 o proteína Cpf1) se prueban para otras proteínas Cas y una matriz espaciadora repetida asociada a Cas se utiliza para identificar todas las secuencias que pertenecen al locus CRISPR-Cas. Esto se puede hacer mediante alineamiento con otros loci CRISPR de Clase 2.

20 Una vez que se identifica la secuencia del locus CRISPR de Clase 2 para el ortólogo de la nucleasa para la especie, se utiliza una exploración predictiva *in silico* para extraer la secuencia de crARN. La secuencia de crARN está contenida en la matriz repetida CRISPR y se pueden identificar por sus secuencias repetidas distintivas interespariadas por secuencias espaciadoras ajenas.

25 B. Preparación de la biblioteca RNA-Seq

La supuesta matriz CRISPR que contiene el crARN identificado *in silico* se valida adicionalmente utilizando secuenciación de ARN (RNA-seq).

30 Las células de las especies identificadas como que comprenden el supuesto crARN se procuraron de un distribuidor comercial (por ejemplo, ATCC, Manassas, VA; German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), Braunschweig, Alemania).

35 Las células se cultivan hasta la fase log media y se preparó el ARN total utilizando el reactivo Trizol (Sigma Aldrich, St. Louis, MO) y se trató con DNasa I (Fermentas, Vilnius, Lituania).

Se trataron 10 µg de ARN total con el Kit Ribo-Zero rRNA Removal (Illumina, San Diego, CA) y el resto de ARN purificado utilizando RNA Clean and Concentrators (Zymo Research, Irvine, CA).

40 Entonces se preparó una biblioteca utilizando el kit de preparación de biblioteca TruSeq Small RNA (Illumina, San Diego, CA) siguiendo las instrucciones del fabricante. Esto da como resultado ADNc que tienen secuencias adaptadoras.

La biblioteca de ADNc resultante se secuenció utilizando el Secuenciador TruSeq (Illumina, San Diego, CA).

45

C. Procesamiento de los datos de secuenciación

Las lecturas de secuenciación de la biblioteca de ADNc se procesan, por ejemplo, utilizando el siguiente método.

50 Las secuencias adaptadoras se retiraron utilizando cutadapt 1.1 ([pypi.python.org/pypi/cutadapt/1.1](http://pypi.python.org/pypi/cutadapt/1.1)) y aproximadamente se recortaron 15 nt del extremo 3' de la lectura para mejorar la calidad de lectura.

Las lecturas se alinearon con el genoma de las especies respectivas (es decir, del cual se identificó el supuesto crARN) con una tolerancia de falta de coincidencia de 2 nucleótidos.

55

La cobertura de lectura se calculó utilizando BedTools ([bedtools.readthedocs.org/en/latest/](http://bedtools.readthedocs.org/en/latest/)).

Se utilizó el Integrative Genomics Viewer (IGV; [www.broadinstitute.org/igv/](http://www.broadinstitute.org/igv/)) para mapear la posición de inicio (5') y el final (3') de las lecturas. Las lecturas totales recuperadas para el supuesto crARN se calcularon a partir del archivo SAM de alineamientos.

60

Los datos RNA-seq se utilizaron para validar que el supuesto elemento crARN se transcribe activamente *in vivo*. Los aciertos confirmados de la comparación en las exploraciones *in silico* y RNA-seq se validaron en cuanto a la capacidad funcional para soportar la escisión por nucleasa CRISPR de Clase 2 de secuencias de ácido nucleico diana de ADN de cadena doble utilizando los métodos perfilados en el presente documento (por ejemplo, en los Ejemplos 1, 2, y 3). Se sabe en la técnica que el sistema de Tipo V solo necesita un crARN para facilitar la escisión por la nucleasa Cpf1

65

de una secuencia de ADN diana de cadena doble, mientras que el sistema Tipo II necesita un crARN y un tracrARN equivalente para facilitar la escisión por nucleasa Cas9 de una secuencia de ADN diana de cadena doble. El tracrARN equivalente se puede identificar siguiendo el método descrito en el Ejemplo 6.

- 5 Siguiendo las directrices de la presente memoria descriptiva y los Ejemplos, la identificación de nuevas secuencias de crARN relacionadas con las proteínas Cas puede ser practicada por un experto habituado en la técnica.

### Ejemplo 6

#### 10 Identificación y exploración de tracrARN

Este Ejemplo describe un método por el que se pueden identificar los tracrARN de especies que tienen un sistema CRISPR-Cas9 de Clase 2 Tipo II. Esto se ha adaptado de Chylinski, K., et al, RNA Biology 10(5):726-37 (2013). No son necesarias las siguientes etapas para la exploración ni se debe seguir el orden de las etapas como se presentan.

#### 15 A. Identificación de especies que contienen un sistema CRISPR-Cas9 Tipo II

Utilizando una Herramienta de búsqueda de alineamiento local básico (BLAST, blast.ncbi.nlm.nih.gov/Blast.cgi), se lleva a cabo una búsqueda de los genomas de distintas especies para identificar una proteína Cas9. Los sistemas CRISPR-Cas9 de Clase 2 Tipo II presentan una alta diversidad de secuencia a través de las especies, sin embargo, los ortólogos de Cas9 tienen dominios conservados, por ejemplo, un dominio endonucleasa HNH y/o un dominio de división RuvC/RNasa H. Los resultados del BLAST primarios se filtran para los dominios identificados; las secuencias incompletas o truncadas se descartan y se identifican los ortólogos de Cas9.

25 Si se identifica un ortólogo de Cas9 en una especie, las secuencias adyacentes a la secuencia codificante del ortólogo de Cas9 se prueban para otras proteínas Cas y una matriz espaciadora repetida asociada a Cas se utiliza para identificar todas las secuencias que pertenecen al locus CRISPR-Cas9. Esto se puede hacer mediante alineamiento con otros loci CRISPR-Cas9 de Clase 2 Tipo II conocidos, con el conocimiento de que especies relacionadas estrechamente presentan una arquitectura de locus CRISPR-Cas9 similar (por ejemplo, la composición, tamaño, orientación, localización de la matriz, localización del tracrARN, y así) El elemento tracrARN está contenido normalmente con el locus CRISPR-Cas9 de Clase 2 Tipo II y se identifica fácilmente por su complementariedad de secuencias con los elementos repetidos en la matriz espaciadora repetida. La complementariedad de secuencias tracr con los elementos repetidos se llama el tracr anti-secuencias repetidas.

35 Una vez que se identifica la secuencia del locus CRISPR-Cas9 correspondiente al ortólogo de Cas9 para una especie, se utiliza una exploración predictiva *in silico* para extraer la secuencia de tracrARN anti-secuencia repetida para identificar el tracrARN asociado a Cas. Las supuestas anti-repeticiones se exploran, por ejemplo, de la siguiente manera.

40 Si la secuencia repetida es de una especie conocida, se identifica y se recupera de la base de datos CRISPRdb (crispr.u-psud.fr/crispr). Si la secuencia repetida no se sabe que esté relacionada con una especie, las secuencias repetidas se prevén empelando el software CRISPRfinder (crispr.u-psud.fr/Server/) utilizando el locus CRISPR-Cas9 de Clase 2 Tipo II para la especie como se ha descrito anteriormente.

45 La secuencia repetida identificada para la especie se utiliza para sondear el locus CRISPR-Cas9 para la secuencia anti-repetición (por ejemplo, utilizando el algoritmo BLASTp o similares). La búsqueda se restringe normalmente a las regiones intergénicas del locus CRISPR-Cas9.

Una región tracr anti-repeticiones se valida por complementariedad con la secuencia repetida identificada.

50 Una supuesta región anti-repetición se sondea en las regiones 5' y 3' de la supuesta región anti-repetición en cuanto a la presencia de un terminador transcripcional independiente de Rho (TransTerm HP, transterm.cbcb.umd.edu).

55 Combinando la secuencia identificada que comprende el elemento anti-repetición y el terminador transcripcional independiente de Rho se determina que la secuencia sea el supuesto tracrARN de la especie determinada.

#### B. Preparación de la biblioteca RNA-Seq

60 El supuesto tracrARN, identificado *in silico*, se valida adicionalmente utilizando la secuenciación de ARN (RNA-seq).

Las células de especies que comprenden el supuesto tracrARN se adquieren en un distribuidor comercial (por ejemplo, ATCC, Manassas VA; DSMZ, Braunschweig, Alemania).

65 Las células se cultivaron hasta la fase log media y se preparó el ARN total utilizando el reactivo Trizol (Sigma Aldrich, St. Louis, MO) y se trató con DNasa I (Fermentas, Vilnius, Lituania).

Se trataron 10 µg de ARN total utilizando Kit Ribo-Zero rRNA Removal (Illumina, San Diego, CA) y el resto de ARN purificado utilizando RNA Clean and Concentrators (Zymo Research, Irvine, CA).

5 Entonces se preparó una biblioteca utilizando el Kit de preparación de biblioteca TruSeq Small RNA (Illumina, San Diego, CA) siguiendo las instrucciones del fabricante. Esto da como resultado ADNc que tienen secuencias adaptadoras.

La biblioteca de ADNc resultante se secuenció utilizando el Secuenciador TruSeq (Illumina, San Diego, CA).

10 C. Procesamiento de los datos de secuenciación

Las lecturas de secuenciación de la biblioteca de ADNc se procesan, por ejemplo, utilizando el siguiente método.

15 Las secuencias adaptadoras se retiraron utilizando cutadapt 1.1 ([pypi.python.org/pypi/cutadapt/1.1](http://pypi.python.org/pypi/cutadapt/1.1)) y aproximadamente se recortaron 15 nt del extremo 3' de la lectura para mejorar la calidad de lectura.

Las lecturas se alinearon con el genoma de las especies respectivas (es decir, del cual se identificó el supuesto crARN) con una tolerancia de falta de coincidencia de 2 nucleótidos.

20 La cobertura de lectura se calculó utilizando BedTools ([bedtools.readthedocs.org/en/latest/](http://bedtools.readthedocs.org/en/latest/)).

Se utilizó el Integrative Genomics Viewer (IGV; [www.broadinstitute.org/igv/](http://www.broadinstitute.org/igv/)) para mapear la posición de inicio (5') y el final (3') de las lecturas. Las lecturas totales recuperadas para el supuesto crARN se calcularon a partir del archivo SAM de alineamientos.

25 Los datos RNA-seq se utilizaron para validar que el supuesto elemento crARN se transcribe activamente *in vivo*. Los aciertos confirmados de la comparación en las exploraciones *in silico* y RNA-seq se validaron en cuanto a la capacidad funcional para soportar la escisión por nucleasa CRISPR de Clase 2 de secuencias de ácido nucleico diana de ADN de cadena doble utilizando los métodos perfilados en el presente documento (por ejemplo, en los Ejemplos 1, 2, y 3).

30 Siguiendo las directrices de la presente memoria descriptiva y los Ejemplos, la identificación de nuevas secuencias de tracrARN relacionadas con las proteínas Cas9 puede llevarse a cabo por un experto habituado en la técnica.

### 35 Ejemplo 7

Ensayo T7E1 para la detección de las modificaciones en la secuencia diana en células eucariotas

Este Ejemplo ilustra el uso de ensayo T7E1 para evaluar y comparar el porcentaje de escisión en la célula de complejos Cas9-Cpf1-NATNA/Cas9&Cpf1 con respecto a las secuencias de ADN diana de cadena doble.

40 A. Transfecciones celulares utilizando componentes polinucleotídicos Cas

45 Los Cas9-Cpf1-NATNA se transfectaron en células HEK293 que expresaban constitutivamente Cas9 de *S. pyogenes* y Cpf1 de *Acidaminococcus sp.* (HEK293-Cas9-Cpf1), utilizando el sistema Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ) y el siguiente protocolo. Los Cas9-Cpf1-NATNA se prepararon en tampón de hibridación (1,25 mM de HEPES, 0,625 mM de MgCl<sub>2</sub>, 9,375 mM KCl a pH 7,5), se incubaron durante 2 minutos a 95 °C, se retiraron del termociclador, se permitió que se equilibraran a temperatura ambiente, y se dispensaron con un volumen final de 5 µl en una placa de 96 pocillos. El medio de cultivo se aspira de las células HEK-Cas9-Cpf1, las células se lavaron una vez con PBS libre de calcio y magnesio, y entonces se tripsinizaron por adición de TrypLE (Life Technologies, Grand Island, NY) seguido por incubación a 37 °C durante 3-5 minutos. Las células tripsinizadas se pipetearon suavemente arriba y abajo para formar una suspensión de células únicas y se añadió a un medio de cultivo DMEM completo compuesto por medio de cultivo DMEM (Life Technologies, Grand Island, NY) q que contenía un 10 % de suero fetal bovino (FBS; Thermo Fisher Scientific, Wilmington, DE) y suplementado con penicilina y estreptomina (Life Technologies, Grand Island, NY).

55 Las células se aglomeraron entonces por centrifugación durante 3 minutos a 200 x g, se aspiró el medio de cultivo, y las células se re-suspendieron en PBS. Se hizo el recuento de células utilizando el Contador celular automático Countess® II (Life Technologies; Grand Island, NY). Se transfirieron  $2,2 \times 10^7$  células a un tubo de microfuga de 1,5 ml y se aglomeraron. El PBS se aspiró y las células se re-suspendieron en solución Nucleofector™ SF (Lonza, Allendale, NJ) hasta una densidad de  $1 \times 10^7$  células/ml. Se añaden entonces 20 µl de suspensión celular a cada pocillo individual que contenía 5 µl de Cas9-Cpf1-NATNA, y todo el volumen de cada pocillo se transfirió a un pocillo de una Placa Nucleocuvette™ de 96 pocillos (Lonza, Allendale, NJ). La placa se cargó en el Nucleofector® 96-well Shuttle (Lonza, Allendale, NJ) y las células se nucleofectaron utilizando el programa 96-CM-130 Nucleofector™ (Lonza, Allendale, NJ). Después de la nucleofección se añadieron 70 µl de medio DMEM completo a cada pocillo y se transfirieron 50 µl de la suspensión celular a una placa de cultivo celular de 96 pocillos revestida de colágeno que contenía 150 µl de medio de cultivo DMEM completo precalentado. La placa se transfirió entonces a una incubadora

de cultivo tisular y se mantuvo a 37 °C en un 5 % de CO<sub>2</sub> durante 48 horas.

B. Generación de la secuencia de ADN diana de cadena doble por el ensayo T7E1

- 5 Se aisló el gADN de las células HEK293-Cas9-Cpfl 48 horas después de la transfección del Cas9-Cpfl-NATNA utilizando 50 µl de solución de extracción de ADN QuickExtract (Epicentre, Madison, WI) por pocillo seguido por una incubación a 37 °C durante 10 minutos, 65 °C durante 6 minutos y 95 °C durante 3 minutos para parar la extracción. El gADN se diluyó entonces con 150 µl de agua y se almacenaron las muestras a -80 °C.
- 10 El ADN para el T7E1 se generó por amplificación por PCR de las secuencias de ADN de cadena doble (por ejemplo, VEGFA e IDO2) del gADN aislado. Las reacciones de PCR se llevaron a cabo utilizando 8 µl de gADN como matriz con polimerasa KAPA HiFi Hot Start y contenía 0,5 U de polimerasa, 1x de tampón de reacción, 0,4 mM de dNTPs y 300 nM de cebadores directos e inverso dirigidos a la secuencia de ADN diana de cadena doble (por ejemplo, Ejemplo 2, Tabla 3; SEQ ID NO. 5 y SEQ ID NO. 6; y SEQ ID NO. 7 y SEQ ID NO. 8) en un volumen total de 25 µl. La secuencia de ADN diana se amplifica utilizando las siguientes condiciones: 95 °C durante 5 minutos, 4 ciclos de 20 segundos a 98 °C, 20 segundos a 70 °C, menos 2 °C/ciclo, 30 segundos a 72 °C, seguido por 30 ciclos de 15 segundos a 98 °C, 20 segundos a 62 °C, 20 segundos a 72 °C, y una extensión final a 72 °C durante 1 minuto.

C. Ensayo T7E1

- 20 Las secuencias de ADN diana de cadena doble amplificadas por PCR se desnaturalizaron a 95 °C durante 10 minutos y se permitió que se rehibridaran por enfriamiento a 25 °C a -0,5 °C/s en un termociclador. El ADN rehibridado se incubó a 0,5 µl de endonucleasa T7 1 en 1x de tampón NEBuffer 2 (New England Biolabs, Ipswich, MA) en un volumen total de 15 µl durante 25 minutos a 27 °C. Las reacciones T7E1 se analizaron utilizando el Sistema Fragment Analyzer™ (Advanced Analytical Technologies, Ames, IA). El sistema Fragment Analyzer™ proporciona la concentración de cada fragmento de escisión y de la secuencia de ADN diana de cadena doble que permanece después de la escisión.

- 30 Los porcentajes de escisión de las secuencias de ADN diana de doble cadena se calcularon a partir de la concentración de cada fragmento de escisión y a secuencia de ADN DIANA de doble cadena que permanece después de que haya tenido lugar la escisión, utilizando la siguiente fórmula:

$$\% \text{ escisión} = \left( 1 - \sqrt{1 - \frac{(\text{frag1} + \text{frag2})}{(\text{frag1} + \text{frag2} + \text{parental})}} \right)$$

ECUACIÓN 1

- 35 En la Ecuación 1, las concentraciones frag1 y frag2 se corresponden con la concentración de fragmentos de la escisión por Cas9 de la secuencia de ADN diana de cadena doble y parental se corresponde con la secuencia de ADN diana de cadena doble que permanece después de que haya tenido lugar la escisión.

- 40 El ensayo T7E1 para la detección de modificaciones en la secuencia diana en células eucariotas proporciona datos para demostrar que los complejos de Cas9-Cpfl-NATNA/proteína Cas9&Cpfl descritos en el presente documento facilitan la escisión *in vivo* específica del sitio mediada por Cas9 y Cpfl de secuencias de ADN diana de cadena doble. Los polinucleótidos sgARN, crARN y/o tracrARN/crARN que tienen la misma secuencia de unión del ADN diana como la Cas9-Cpfl-NATNA también se puede incluir en el ensayo para comparar los porcentajes de escisión específica del sitio mediada por Cas9 y Cpfl entre las construcciones.

- 45 Siguiendo las directrices de la presente memoria descriptiva y los Ejemplos, el ensayo T7E1 descrito en este Ejemplo se puede practicar por un experto en la técnica con otras proteínas CRISPR Cas9 de Tipo II o proteínas CRISPR Cpfl de Tipo V y su Cas-Cpfl-NATNA equivalente.

- 50

**Ejemplo 8**

Sondeo de sitios tolerantes a la modificación en las estructuras de crARN guía de Clase 2 Tipo V

5 Este Ejemplo describe la generación y ensayo de distintas modificaciones de los crARN guía de Clase 2 Tipo V y su idoneidad para su uso en la construcción de NANTA de tipo cruzado.

10 en este Ejemplo, se modificaron las secuencias de armazón de crARN, y el crARN modificado se ensayó con una nucleasa CpfI equivalente para facilitar la identificación de regiones o posiciones de la estructura CpfI-crARN en el que se puede unir un Cas9-sgARN mediante enlaces de hidrógeno de pares de bases o un enlace covalente (por ejemplo, enlace fosfodiéster).

15 Se seleccionaron distintas regiones en el CpfI-crARN de *F. novicida* para la modificación. Se encargaron los cebadores para la generación de CpfI-crARN modificado en un fabricante comercial. Las estructuras de ADN de cadena doble para la transcripción de crARN se generaron como se ha descrito previamente (por ejemplo, el Ejemplo 1). Las secuencias utilizadas para el ensamblaje del armazón de crARN se muestran en la Tabla 7.

Tabla 7

Cebadores para el ensamblaje del armazón de crARN		
Espaciador de secuencia diana	Modificación de secuencia de crARN	Cebadores utilizados para el ensamblaje por PCR
VEGFA	crARN de tipo silvestre	SEQ ID NO. 1, 17, 27
IDO2	crARN de tipo silvestre	SEQ ID NO. 1, 18, 4
VEGFA	Cola 5' de cadena sencilla	SEQ ID NO. 1, 19, 27
IDO2	Cola 5' de cadena sencilla	SEQ ID NO. 1, 20, 4
VEGFA	Aumento de longitud de pseudo-nudo	SEQ ID NO. 1, 21, 27
IDO2	Aumento de longitud de pseudo-nudo	SEQ ID NO. 1, 22, 4
VEGFA	Adición de horquilla 3'	SEQ ID NO. 1, 23, 27
IDO2	Adición de horquilla 3'	SEQ ID NO. 1, 24, 4
VEGFA	Adición de horquilla 5'	SEQ ID NO. 1, 25, 27
IDO2	Adición de horquilla 5'	SEQ ID NO. 1, 26, 4

20 La FIG. 14A ilustra el CpfI-crARN de tipo silvestre. La FIG. 14B ilustra el CpfI pre-crARN. Se representan los ejemplos de modificación realizadas en el CpfI-crARN en la FIG. 14C, FIG. 14D, y FIG. 14E, en el que las modificaciones se indican por las partes no sombreadas del CpfI-crARN modificado que se representa. Las secuencias correspondientes del CpfI-crARN se presentan en la Tabla 8, incluyendo el espaciador, tipo de modificación, secuencia de crARN (modificado), y la SEQ ID NO. para la secuencia de crARN (modificada), así como la figura correspondiente a la estructura general para cada crARN (modificado).

25

Tabla 8

Secuencias de CpfI-crARN modificadas				
Espaciador de secuencias diana	Modificación de secuencia de crARN	Secuencia de crARN modificada	SEQ ID NO.	FIG.
VEGFA	crARN de tipo silvestre	GAAUUUCUACUGUUGUAGA UGAAAGGGGGUGGGGGGAG UUJGCU	SEQ ID NO. 28	14A
IDO2	crARN de tipo silvestre	GAAUUUCUACUGUUGUAGA UGGUAGAGAAAGAAGCAGU GCCUGG	SEQ ID NO. 29	14A

VEGFA	Cola 5' de cadena sencilla	GGUCUAAGAACUUUAAAUA AUUUCUACUGUUGUAGAUG AAAGGGGGUGGGGGGAGUU UGCU	SEQ ID NO. 30	14B
IDO2	Cola 5' de cadena sencilla	GGUCUAAGAACUUUAAAUA AUUUCUACUGUUGUAGAUG GUAGAGAAAGAAGCAGUGC CUGG	SEQ ID NO. 31	14B
VEGFA	Aumento de longitud de pseudo-nudo	GAAUUUCUACUGUUCACUG CCGUUAGGCAGUGAACUU GUAGAUGAAAGGGGGUGGG GGGAGUUUGCU	SEQ ID NO. 32	14C
IDO2	Aumento de longitud de pseudo-nudo	GAAUUUCUACUGUUCACUG CCGUUAGGCAGUGAACUU GUAGAUGGUAGAGAAAGAA GCAGUGCCUGG	SEQ ID NO. 33	14C
VEGFA	Adición de horquilla 3'	GAAUUUCUACUGUUGUAGA UGAAAGGGGGUGGGGGGAG UUUGCUGCCAGGUUCACUG CCGUUAGGCAG	SEQ ID NO. 34	14D
IDO2	Adición de horquilla 3'	GAAUUUCUACUGUUGUAGA UGGUAGAGAAAGAAGCAGU GCCUGGGCCAGGUUCACUGC CGUUAAGGCAG	SEQ ID NO. 35	14D
VEGFA	Adición de horquilla 5'	GGUUCACUGCCGUUAGGC AGGCAAGAAUUUCUACUGU UGUAGAUGAAAGGGGGUGG GGGAGUUUGCU	SEQ ID NO. 36	14E
IDO2	Adición de horquilla 5'	GGUUCACUGCCGUUAGGC AGGCAAGAAUUUCUACUGU UGUAGAUGGUAGAGAAAGA AGCAGUGCCUGG	SEQ ID NO. 37	14E

5 La proteína Cas9 de *S. pyogenes* y la proteína Cpf1 de *F. novicida* se expresaron recombinantemente en *E. coli* y se purificaron para su uso en los ensayos de escisión bioquímica *in vitro*. Se generaron las secuencias de ADN diana de carne doble de VEGFA e IDO2 como se ha descrito en el Ejemplo 2. La escisión bioquímica de las secuencias diana generadas por PCR se llevaron a cabo esencialmente como se ha descrito en el Ejemplo 3, con la modificación de que solamente se utilizó la Cpf1 de *F. novicida*.

10 La FIG. 15 presenta los resultados del ensayo de escisión bioquímica de Cpf1 de *F. novicida* utilizando los Cpf1-crARN modificados (Tabla 8). El eje y muestra el porcentaje de escisión de una secuencia de ADN diana de cadena doble, y el eje x muestra la SEQ ID NO. correspondiente al Cpf1-crARN (modificado) que se utiliza en el ensayo. como se puede ver en los datos, se toleraba la inclusión de bases y estructuras secundarias de ARN (por ejemplo, elementos ahorquillados) en el Cpf1-crARN.

15 Los métodos descritos en este Ejemplo se pueden utilizar para identificar localizaciones en una secuencia de Cpf1-crARN que permiten la modificación del Cpf1-crARN por unión a un Cas-crARN, tracrARN, o sgARN para generar un Cas9-Cpf1-NATNA. El Cpf1-crARN de diferentes especies puede presentar diferentes grados de tolerancia para distintas modificaciones de secuencia. Adicionalmente, puede haber diferentes Cas9-crARN, tracrARN o sgARN más dispuestos a la inclusión en un Cas9-Cpf1-NATNA.

20 Los Cpf1-crARN que son capaces de mediar la escisión de una secuencia de ADN diana con sus proteínas Cpf1 equivalentes se pueden validar para la actividad en células utilizando el método descrito en el Ejemplo 4.

25 Este Ejemplo se puede practicar por un experto habituado en la técnica con otras proteínas con CRISPR de Clase 2 Tipo V y su crARN guía con CRISPR Tipo V. Siguiendo las directrices de la presente memoria descriptiva y Ejemplos,

la modificación de un CpfI-crARN (por ejemplo, la introducción de distintas modificaciones de secuencia y estructura secundaria, y/o eliminación de distintas secuencias) se pueden utilizar para sondear localizaciones para la inserción o uniones para facilitar la producción de un Cas9-CpfI-NATNA.

## 5 Ejemplo 9

### Sondeo de sitios tolerantes a la modificación en estructuras de ARN guía Cas9 de Clase 2 Tipo II

Este ejemplo describe la generación y ensayo de distintas modificaciones de ARN guía de Clase 2 Tipo II y su idoneidad para su uso en la construcción de NATNA de tipo cruzado.

En este Ejemplo, se introducen modificaciones en la estructura del ARN de los ARN guía con CRISPR de Clase 2 Tipo II para identificar localizaciones para modificar o unir distintas secuencias de ácido nucleico. El método que se describe posteriormente está adaptado de Briner, A., et al, *Molecular Cell* 56(2):333-339 (2014). No son necesarias todas las etapas siguientes para la exploración ni se debe seguir el orden de las etapas como se presentan.

Se seleccionó un sgARN de un sistema CRISPR de Clase 2 Tipo II para la modificación. La secuencia de ARN guía se modificó *in silico* para introducir sustituciones de bases en las regiones seleccionadas a partir de uno o más de los siguientes, un tallo inferior, protuberancia, tallo superior, tetrabucle, nexos, secuencias de unión, y horquillas 3'. La sustitución de bases se puede utilizar para introducir faltas de coincidencia en las interacciones de hidrógeno de pares de base de cualquiera de las regiones del ARN guía, o la mutación puede conservar la interacción de hidrógeno de pares de bases mediante la sustitución de dos bases que introducen pares de bases de Watson-Crick alternativas. La sustitución de bases también se puede utilizar para introducir nuevas interacciones de hidrógeno de pares de bases en el almacén del ARN guía (por ejemplo, en la secuencia de la protuberancia).

Se pueden modificar independientemente regiones del ARN guía para introducir elementos de estructura secundaria en el almacén de del ARN guía. Dichas modificaciones incluyen horquillas, pseudo-nudos, ribozimas, u o tras modificaciones conocidas por el experto en la técnica. Además, el almacén de ARN guía se puede modificar para borrar partes del almacén del ARN guía, sea mediante eliminación en el extremo 5', el extremo 3' o internos en el ARN guía.

Las secuencias de ARN guía CRISPR de Clase 2 Tipo II diseñadas *in silico* se proporcionaron a un fabricante comercial para las síntesis.

Los ARN guía CRISPR de Clase 2 Tipo II se evaluaron en cuanto a su capacidad para soportar la escisión de una secuencia de ADN diana de cadena doble mediada por sus proteínas Cas9 equivalentes. Las secuencias de ADN diana de cadena doble y el ensayo de escisión bioquímica se llevaron a cabo de manera similar a las descritas en el Ejemplo 2 y Ejemplo 3. Los ARN guía que son capaces de mediar en la escisión de la secuencia de ADN diana con sus proteínas Cas9 equivalentes se pueden validar en cuanto a la actividad en células utilizando el método descrito en el Ejemplo 4.

Siguiendo las directrices de la presente memoria descriptiva y Ejemplos, la modificación de ARN guía Cas9 (por ejemplo, la introducción de distintas modificaciones de secuencia y estructura secundaria, y/o la eliminación de distintas secuencias) se puede utilizar para sondear localizaciones para la inserción o uniones para facilitar la producción de un Cas9-CpfI-NATNA. Este Ejemplo puede practicarse por un experto en la técnica con otras proteínas Cas9 CRISPR Tipo II y otro ARN guía CRISPR Tipo II.

## Ejemplo 10

### Exploración de NATNA de tipo cruzado que comprenden secuencias de unión a ADN diana

Este Ejemplo ilustra el uso de NATNA de tipo cruzado de la presente invención para modificar secuencias de ADN diana presentes en el ADN genómico humano y para medir el nivel de actividad de escisión en esos sitios.

Los sitios diana (secuencias de ADN diana) se seleccionan primero a partir del ADN genómico. Se diseñan Cas9-CpfI-NATNA para dirigirse a las secuencias seleccionadas. Se llevan a cabo ensayos (por ejemplo, como se describe en el Ejemplo 3) para determinar el nivel de escisión de la secuencia de ADN diana.

No son necesarios todas las etapas siguientes para cada exploración ni se debe seguir el orden de las etapas como se presentan, y la exploración se puede acoplar con otros experimentos, o formar parte de un experimento mayor.

A. Selección de regiones diana (secuencias de ADN diana) del ADN genómico

Todas las secuencias PAM (por ejemplo, NGG, TTTN) se identifica en la región genómica seleccionada.

65

Se identifican y seleccionan una o más de entre las 20 secuencias de nucleótidos (secuencias de ADN diana Cas9) que estén adyacentes 5' a una secuencia PAM NGG.

5 Se identifican y seleccionan una o más de entre las 24 secuencias de nucleótidos (secuencias de ADN diana Cpfl) que están adyacentes 3' a una secuencia PAM TTTN.

10 Los criterios de selección de secuencias de ácido nucleico diana pueden incluir, pero no se limitan a, los siguientes: homología con otras regiones del genoma; porcentaje del contenido G-C; temperatura de fusión; presencias de homopolímero en el espaciador; distancia entre las dos secuencias; y otros criterios conocidos por el experto en la técnica.

15 La secuencia de unión de ADN Cas9 diana está adjunta al extremo 5' de un Cas9-Cpfl-NATNA apropiado, y la secuencia de unión del Cpfl está adjunta al extremo 3' del mismo Cas-Cpfl-NATNA (por ejemplo, un Cas9-Cpfl-NATNA como se ilustra en la FIG. 4). Una construcción Cas9-Cpfl-NATNA se sintetiza normalmente por un fabricante comercial, y se basa en las secuencias proporcionadas al fabricante. De manera alternativa, la construcción Cas9-Cpfl-NATNA se produce como se ha descrito en el Ejemplo 1 por transcripción *in vitro*.

20 Un Cas9-Cpfl-NATNA, como se describe, en el presente documento, se utiliza con las proteínas Cas CRISPR de Clase 2 Tipo II y Clase 2 Tipo V (por ejemplo, ácidos nucleicos asociados con Cas9 de un Cas9-Cpfl-NATNA unido a su proteína Cas9 equivalente, y los ácidos nucleicos asociados con Cpfl de un Cas9-Cpfl-NATNA unido a su proteína Cpfl equivalente).

#### B. Determinación de los porcentajes y especificidad de escisión

25 Los porcentajes y especificidad de la escisión *in vitro* (por ejemplo, la cantidad de unión fuera de la diana) con respecto al Cas9-Cpfl-NATNA se determinaron, por ejemplo, utilizando los ensayos de escisión descritos en el Ejemplo 3, y se compararon de la siguiente manera:

30 (1) Si solo se identifica o selecciona un único par de secuencias de ADN diana para un Cas9-Cpfl-NATNA, se determina el porcentaje y especificidad de escisión de cada una de las secuencias de ADN diana. Si se desea, el porcentaje y/o especificidad de escisión se alteran en experimentos adicionales utilizando métodos que incluyen, pero no se limitan a, modificación del Cas9-Cpfl-NATNA, o introducción de proteínas efectoras/ secuencias de unión a proteínas efectoras para modificar el Cas9-Cpfl-NATNA o una o ambas de las proteínas Cas, o restos de unión ligando/ligando para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas Cas.

35 (2) Si se identifican o seleccionan múltiples pares de secuencias de ADN diana para un Cas9-Cpfl-NATNA, los datos del porcentaje de escisión y datos de especificidad del sitio obtenidos de los ensayos de escisión se comparan entre diferentes ADN que comprenden secuencias de unión a la diana para identificar las secuencias de ADN diana que tengan el porcentaje y especificidad de escisión deseados. Los datos de porcentaje de escisión y datos de especificidad proporcionan los criterios de elección de bases para varias aplicaciones. Por ejemplo, en algunas situaciones la actividad de los Cas9-Cpfl-NATNA puede ser el factor más importante. En otras situaciones, la especificidad del sitio de escisión puede ser relativamente más importante que el porcentaje de escisión. Si se desea, el porcentaje y/o especificidad de escisión se alteran en experimentos adicionales utilizando métodos de la presente invención incluyendo, pero sin limitarse a, modificación del Cas9-Cpfl-NATNA, o introducción de proteínas efectoras/secuencias de unión a proteína efectora para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas Cas, o restos de unión ligando/ligando para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas Cas.

50 De manera alternativa o además de los análisis *in vitro*, se compararon los porcentajes y especificidades de escisión en células asociados con Cas9-Cpfl-NATNA obtenidos utilizando, por ejemplo, el método descrito en el Ejemplo 5, de la siguiente manera:

(1) Si solo se identifica o selecciona un único par de secuencias de ADN diana para un Cas9-Cpfl-NATNA, se determina el porcentaje y especificidad de escisión de cada una de las secuencias de ADN diana. Si se desea, el porcentaje y/o especificidad de escisión se alteran en experimentos adicionales utilizando los que incluyen, pero no se limitan a, modificación del Cas9-Cpfl-NATNA, o introducción de proteínas efectoras/ secuencias de unión a proteínas efectoras para modificar el Cas9-Cpfl-NATNA o una o ambas de las proteínas Cas, o restos de unión ligando/ligando para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas Cas.

60 (2) Si se identifican o seleccionan múltiples pares de secuencias de ADN diana para un Cas9-Cpfl-NATNA, los datos del porcentaje de escisión y datos de especificidad del sitio obtenidos de los ensayos de escisión se comparan entre diferentes ADN que comprenden secuencias de unión a la diana para identificar las secuencias de ADN diana que tengan el porcentaje y especificidad de escisión deseados. Los datos de porcentaje de escisión y datos de especificidad proporcionan los criterios de elección de bases para varias aplicaciones. Por ejemplo, en algunas situaciones la actividad de los Cas9-Cpfl-NATNA puede ser el factor más importante. En otras situaciones, la especificidad del sitio de escisión puede ser relativamente más importante que el porcentaje de escisión. Si se desea, el porcentaje y/o especificidad de escisión se alteran en experimentos adicionales utilizando métodos de la presente invención incluyendo, pero sin limitarse a, modificación del Cas9-Cpfl-NATNA, o introducción de proteínas efectoras/secuencias de unión a proteína efectora para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas

Cas, o restos de unión ligando/ligando para modificar el Cas9-Cpfl-NATNA o una o ambas proteínas Cas.

5 Siguiendo las directrices de la presente memoria descriptiva y Ejemplo, la exploración descrita en este Ejemplo puede practicarse por un experto en la técnica con otros NATNA de tipo cruzado para su uso con proteínas Cas9 CRISPR de Clase 2 Tipo II y proteínas Cpfl CRISPR de Clase 2 Tipo V.

LISTADO DE SECUENCIAS

10	<110> Caribou Biosciences, Inc.	
	<120> Ácido nucleico modificado que se dirige a ácidos nucleicos	
	<130> CBI019.30	
15	<150> 62/245.918	
	<151> 23-10-2015	
	<160> 37	
20	<170> PatentIn versión 3.5	
	<210> 1	
	<211> 22	
	<212> ADN	
25	<213> Secuencia artificial	
	<220>	
	<223> Secuencia de cebador	
30	<400> 1	
	<b>agtgtaatac gactcactat ag</b>	<b>22</b>
	<210> 2	
	<211> 91	
35	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Secuencia de cebador	
40	<400> 2	
	<b>taatacgact cactataggg gtggggggag tttgctccgt ttagagcta gaaatagcaa</b>	<b>60</b>
	<b>gttgagataa ggctagtccg ttatcaactt g</b>	<b>91</b>
	<210> 3	
45	<211> 91	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
50	<223> Secuencia de cebador	
	<400> 3	
	<b>ccaggcactg cttctttctc taccatctac aacagtagaa attaagcacc gactcgggtc</b>	<b>60</b>
	<b>cactttttca agttgataac ggactagcct t</b>	<b>91</b>
55	<210> 4	
	<211> 24	
	<212> ADN	
	<213> Secuencia artificial	
60	<220>	

ES 2 699 848 T3

<223> Secuencia de cebador

<400> 4  
ccaggcactg cttcttctc tacc 24

5

<210> 5  
<211> 141  
<212> ARN  
<213> Secuencia artificial

10

<220>  
<223> Secuencia Cas9-CpfI NATNA1

<400> 5  
gggugggggg aguuugcucc guuuuagagc uagaaauagc aaguugagau aaggcuaguc 60  
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuaa uuucuacugu uguagauggu 120  
agagaaagaa gcagugccug g 141

15

<210> 6  
<211> 52  
<212> ADN  
<213> Secuencia artificial

20

<220>  
<223> Secuencia de cebador

25

<400> 6  
cactcttcc ctacacgacg ctctccgat ctccagatgg cacattgca ga 52

<210> 7  
<211> 47  
<212> ADN  
<213> Secuencia artificial

30

<220>  
<223> Secuencia de cebador

35

<400> 7  
ggagttcaga cgtgtgctct tccgatctcc tagtgactgc cgtctgc 47

<210> 8  
<211> 52  
<212> ADN  
<213> Secuencia artificial

40

<220>  
<223> Secuencia de cebador

45

<400> 8  
cactcttcc ctacacgacg ctctccgat ctaaggaacc tggagacat ca 52

<210> 9  
<211> 48  
<212> ADN  
<213> Secuencia artificial

50

<220>  
<223> Secuencia de cebador

55

<400> 9  
ggagttcaga cgtgtgctct tccgatctga aaggcactga gtgggaag 48

60

ES 2 699 848 T3

<210> 10  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <223> Secuencia de Protoespaciador  
 <400> 10  
 gTgcCctcgg Tcacactgtg Tgg 23  
 10  
 <210> 11  
 <211> 28  
 <212> ADN  
 <213> Secuencia artificial  
 15  
 <220>  
 <223> Secuencia de Protoespaciador  
 <400> 11  
 Tttctctgat gTtcCctcaa ctgtcacc 28  
 20  
 <210> 12  
 <211> 141  
 <212> ARN  
 <213> Secuencia artificial  
 25  
 <220>  
 <223> Secuencia de ERBB2-Cas9-Cpfl NATNA  
 <400> 12  
 gUgcccucgg ucacacugug guuuuagagc uagaaaauagc aaguugagau aaggcuaguc 60  
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuaa uuucuacugu uguagauucu 120  
 gauguucccu caacugucac c 141  
 30  
 <210> 13  
 <211> 47  
 <212> ADN  
 <213> Secuencia artificial  
 35  
 <220>  
 <223> Secuencia de cebador  
 <400> 13  
 cactcttcc ctacacgagc ctctccgat ctctgccccg ggcgctg 47  
 40  
 <210> 14  
 <211> 47  
 <212> ADN  
 <213> Secuencia artificial  
 45  
 <220>  
 <223> Secuencia de cebador  
 <400> 14  
 ggagttcaga cgtgtgctct tccgatctgg tcacaggggt tggaaagg 47  
 50  
 <210> 15  
 <211> 57  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 60

ES 2 699 848 T3

<220>  
 <223> Secuencia de cebador  
  
 <400> 15  
 5 caagcagaag acggcatacg agattacgtg atgtgactgg agttcagacg tgtgctc 57  
  
 <210> 16  
 <211> 58  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de cebador  
  
 15 <400> 16  
 aatgatacgg cgaccacoga gatctacacc gtctaataca ctcttccct acacgacg 58  
  
 <210> 17  
 <211> 61  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de cebador  
 25  
 <400> 17  
**taatacgact cactatagaa tttctactgt tgtagatgaa agggggtggg gggagtttgc 60**  
**t 61**  
  
 <210> 18  
 <211> 61  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de cebador  
 35  
 <400> 18  
**taatacgact cactatagaa tttctactgt tgtagatggt agagaaagaa gcagtgcctg 60**  
**g 61**  
  
 <210> 19  
 <211> 78  
 <212> ADN  
 <213> Secuencia artificial  
 40  
  
 <220>  
 <223> Secuencia de cebador  
 45  
 <400> 19  
**taatacgact cactataggt ctaagaactt taaataattt ctactgttgt agatgaaagg 60**  
**gggtgggggg agtttgc 78**  
 50  
  
 <210> 20  
 <211> 78  
 <212> ADN  
 <213> Secuencia artificial  
 55  
  
 <220>  
 <223> Secuencia de cebador

ES 2 699 848 T3

	<400> 20		
		<b>taatacgact cactataggt ctaagaactt taaataattt ctactgttgt agatggtaga</b>	<b>60</b>
		<b>gaaagaagca gtgcctgg</b>	<b>78</b>
5	<210> 21 <211> 85 <212> ADN <213> Secuencia artificial		
10	<220> <223> Secuencia de cebador		
	<400> 21		
		<b>taatacgact cactatagaa tttctactgt tcaactgcogt ataggcagtg aacttgtaga</b>	<b>60</b>
		<b>tgaaaggggg tggggggagt ttgct</b>	<b>85</b>
15	<210> 22 <211> 85 <212> ADN <213> Secuencia artificial		
20	<220> <223> Secuencia de cebador		
	<400> 22		
		<b>taatacgact cactatagaa tttctactgt tcaactgcogt ataggcagtg aacttgtaga</b>	<b>60</b>
25		<b>tggtagagaa agaagcagtg cctgg</b>	<b>85</b>
30	<210> 23 <211> 86 <212> ADN <213> Secuencia artificial		
	<220> <223> Secuencia de cebador		
35	<400> 23		
		<b>taatacgact cactatagaa tttctactgt tgtagatgaa aggggggtggg gggagtttgc</b>	<b>60</b>
		<b>tgccaggttc actgcogtat aggcag</b>	<b>86</b>
40	<210> 24 <211> 86 <212> ADN <213> Secuencia artificial		
	<220> <223> Secuencia de cebador		
45	<400> 24		
		<b>taatacgact cactatagaa tttctactgt tgtagatggt agagaaagaa gcagtgcctg</b>	<b>60</b>
		<b>ggccaggttc actgccgtat aggcag</b>	<b>86</b>
50	<210> 25 <211> 86 <212> ADN <213> Secuencia artificial		

<220>  
 <223> Secuencia de cebador

5 <400> 25  
**taatacgact cactataggt tcaactgccgt ataggcaggc aagaatttct actgttgtag 60**  
**atgaaagggg gtggggggag tttgct 86**

<210> 26  
 <211> 86  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> Secuencia de cebador

15 <400> 26  
**taatacgact cactataggt tcaactgccgt ataggcaggc aagaatttct actgttgtag 60**  
**atggtagaga aagaagcagt gcctgg 86**

20 <210> 27  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

25 <220>  
 <223> Secuencia de cebador

<400> 27  
 agcaaaactcc cccaccccc tttc 24

30 <210> 28  
 <211> 44  
 <212> ARN  
 <213> Secuencia artificial

35 <220>  
 <223> Secuencia de Cpfl-crARN modificado

<400> 28  
 gaauuucuaC uguuguagau gaaagggggg ggggggaguu ugcu 44

40 <210> 29  
 <211> 44  
 <212> ARN  
 <213> Secuencia artificial

45 <220>  
 <223> Secuencia de Cpfl-crARN modificado

<400> 29  
 gaauuucuaC uguuguagau gguagagaaa gaagcagugc cugg 44

50 <210> 30  
 <211> 61  
 <212> ARN  
 <213> Secuencia artificial

55 <220>  
 <223> Secuencia de Cpfl-crARN modificado

ES 2 699 848 T3

	<400> 30		
	<b>ggucuaagaa cuuuaaaauaa uuucuacugu uguagaugaa aggggguggg gggaguugc</b>	60	
	<b>u</b>	61	
5	<210> 31 <211> 61 <212> ARN <213> Secuencia artificial		
10	<220> <223> Secuencia de Cpfl-crARN modificado		
	<400> 31		
	<b>ggucuaagaa cuuuaaaauaa uuucuacugu uguagauggu agagaaagaa gcagugccug</b>	60	
	<b>g</b>	61	
15	<210> 32 <211> 68 <212> ARN <213> Secuencia artificial		
20	<220> <223> Secuencia de Cpfl-crARN modificado		
	<400> 32		
	<b>gaauuucuaac uguucacugc cguauaggca gugaacuugu agaugaaagg gggugggggg</b>	60	
	<b>aguuugcu</b>	68	
25	<210> 33 <211> 68 <212> ARN <213> Secuencia artificial		
30	<220> <223> Secuencia de Cpfl-crARN modificado		
35	<400> 33		
	<b>gaauuucuaac uguucacugc cguauaggca gugaacuugu agaugguaga gaaagaagca</b>	60	
	<b>gugccugg</b>	68	
40	<210> 34 <211> 69 <212> ARN <213> Secuencia artificial		
45	<220> <223> Secuencia de Cpfl-crARN modificado		
	<400> 34		
	<b>gaauuucuaac uguuguagau gaaagggggg ggggggaguu ugcugccagg uucacugccg</b>	60	
	<b>uauaggcag</b>	69	
50	<210> 35 <211> 69 <212> ARN <213> Secuencia artificial		

# ES 2 699 848 T3

	<220>		
	<223> Secuencia de Cpfl-crARN modificado		
5	<400> 35		
	<b>gaauuucuac uguuguagau gguagagaaa gaagcagugc cugggccagg uucacugccg</b>		<b>60</b>
	<b>uauaggcag</b>		<b>69</b>
	<210> 36		
	<211> 69		
10	<212> ARN		
	<213> Secuencia artificial		
	<220>		
	<223> Secuencia de Cpfl-crARN modificado		
15	<400> 36		
	<b>gguucacugc cguauaggca ggcaagaauu ucuacuguug uagaugaaag gggguggggg</b>		<b>60</b>
	<b>gaguugcu</b>		<b>69</b>
	<210> 37		
	<211> 69		
20	<212> ARN		
	<213> Secuencia artificial		
	<220>		
	<223> Secuencia de Cpfl-crARN modificado		
25	<400> 37		
	<b>gguucacugc cguauaggca ggcaagaauu ucuacuguug uagaugguag agaaagaagc</b>		<b>60</b>
	<b>agugccugg</b>		<b>69</b>

## REIVINDICACIONES

1. Un ácido nucleico de tipo cruzado con CRISPR de Clase 2 modificado que se dirige a ácidos nucleicos ("NATNA de tipo cruzado con CRISPR de Clase 2"), que comprende:

- 5 un ácido nucleico asociado a Cpf1 que se dirige a ácidos nucleicos, que tiene un extremo 5' y un extremo 3', que comprende un elemento espaciador ("Cpf1-NATNA");  
 un primer ácido nucleico asociado a Cas9 que se dirige a ácidos nucleicos, que tiene un extremo 5' y un extremo 3', que comprende un elemento espaciador ("primer Cas9-NATNA"); y  
 10 un segundo ácido nucleico asociado a Cas9 que se dirige a ácidos nucleicos, que tiene un extremo 5' y un extremo 3', que comprende un elemento tracr ("segundo Cas9-NATNA");  
 en donde el primer Cas9-NATNA o el segundo Cas9-NATNA están conectados a Cpf1-NATNA.

2. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de la reivindicación 1 en el que el primer Cas9-NATNA o el segundo Cas9-NATNA:

- 15 están conectados no covalentemente al Cpf1-NATNA mediante enlaces de hidrógeno de pares de bases en el extremo 5' o el extremo 3' del primer Cas9-NATNA o mediante enlaces de hidrógeno de pares de bases en el extremo 5' o el extremo 3' del segundo Cas9-NATNA, o  
 20 están conectados covalentemente al Cpf1-NATNA; y  
 en donde el Cpf1-NATNA está conectado covalentemente al extremo 5' o al extremo 3' del primer Cas9-NATNA o con el extremo 5' o el extremo 3' del segundo Cas9-NATNA.

3. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de las reivindicaciones 1 o 2, en el que el Cpf1-NATNA es capaz de formar un primer complejo con la proteína Cpf1 ("complejo Cas9-Cpf1-NATNA/proteína Cpf1"); y en donde el primer Cas9-NATNA y el segundo Cas9-NATNA son capaces de formar un segundo complejo con una proteína Cas9 ("complejo Cas9-Cpf1-NATNA/proteína Cas9").

4. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de la reivindicación 3,  
 30 en el que, si el Cpf1-NATNA forma el primer complejo con la proteína Cpf1 ("complejo Cpf1-NATNA/proteína Cpf1"), el primer complejo es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble complementaria del elemento espaciador de Cpf1; y en donde si el primer Cas9-NATNA y el segundo Cas9-NATNA forman el segundo complejo con la proteína Cas9 ("complejo Cas9-NATNA/proteína Cas9"), el segundo complejo es capaz de unirse a una segunda secuencia de ácido nucleico diana de cadena doble complementaria del primer elemento espaciador de  
 35 Cas9-NATNA, o  
 en donde, si está formado el primer complejo y está formado el segundo complejo, dando como resultado un complejo Cas9-Cpf1-NATNA/proteínas Cas9&Cpf1, el complejo Cas9-Cpf1-NATNA/proteínas Cas9&Cpf1 es capaz de unirse a una primera secuencia de ácido nucleico diana de cadena doble complementaria del elemento espaciador de Cpf1 y una segunda secuencia de ácido nucleico diana de cadena doble complementaria del primer elemento espaciador de  
 40 Cas9-NATNA.

5. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de cualquier reivindicación anterior,  
 en el que el Cpf1-NATNA comprende adicionalmente una secuencia de nucleótidos de un elemento enlazador conectada covalentemente al extremo 5' o al extremo 3' del Cpf1-NATNA y/o  
 45 en el que el primer Cas9-NATNA comprende adicionalmente una secuencia de nucleótidos de elemento enlazador conectada covalentemente al extremo 5' o al extremo 3' del primer Cas9-NATNA, o el segundo Cas9-NATNA comprende adicionalmente una secuencia de nucleótidos de elemento enlazador conectada covalentemente al extremo 5' o al extremo 3' del segundo Cas9-NATNA.

6. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de la reivindicación 1, en el que el extremo 3' del primer Cas9-NATNA está conectado covalentemente mediante un elemento de bucle al extremo 5' del segundo Cas9-NATNA, dando como resultado un ácido nucleico asociado con Cas9 sencillo que se dirige a ácidos nucleicos ("Cas9-NATNA sencillo"), que tiene un extremo 5' y un extremo 3'; y en donde el Cas9-NATNA sencillo comprende el primer Cas9-NATNA y el segundo Cas9-NATNA.

7. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de la reivindicación 6, en el que el Cas9-NATNA sencillo  
 55 está conectado no covalentemente al Cpf1-NATNA mediante enlaces de hidrógeno de pares de bases en el extremo 5' o el extremo 3' del Cas9-NATNA sencillo, o  
 60 está conectado covalentemente al Cpf1-NATNA; y en donde el Cpf1-NATNA está conectado covalentemente al extremo 5' o al extremo 3' del Cas9-NATNA sencillo.

8. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de una cualquiera de las reivindicaciones 6 o 7,  
 en el que el Cas9-NATNA sencillo comprende una secuencia de nucleótidos de elemento enlazador conectada covalentemente al extremo 5' o al extremo 3' del Cas9-NATNA sencillo y/o  
 65 en el que el Cpf1-NATNA comprende una secuencia de nucleótidos de elemento enlazador conectada covalentemente

al extremo 5' o al extremo 3' del Cpfl-NATNA.

9. El NATNA de tipo cruzado con CRISPR de Clase 2 modificado de cualquier reivindicación anterior, en el que al menos uno de entre el Cpfl-NATNA, el primer Cas9-NATNA o el segundo Cas9-NATNA comprende ARN y/o ADN.

5 10. Una composición de ácido nucleico/proteína, que comprende:

10 el NATNA de tipo cruzado con CRISPR de Clase 2 modificado de cualquiera de las reivindicaciones anteriores, una proteína Cas9, y una proteína Cpfl, en donde el NATNA de tipo cruzado con CRISPR de Clase 2 modificado está formando opcionalmente un complejo con la proteína Cas9 y la proteína Cpfl.

15 11. La composición ácido nucleico/proteína de la reivindicación 10, en la que la proteína Cpfl, la proteína Cas9, o las proteínas Cpfl y Cas9 son enzimáticamente inactivas, o en la que la proteína Cpfl es enzimáticamente inactiva (proteína dCpfl) o la proteína Cas9 es enzimáticamente inactiva (proteína dCas9); y en la que la composición de ácido nucleico/proteína comprende adicionalmente un polinucleótido donante conectado no covalentemente a la proteína dCpfl o la proteína dCas9.

20 12. Un vector de expresión, que comprende: una o más secuencias de ácido nucleico que codifican el NATNA de tipo cruzado con CRISPR de Clase 2 modificado de una cualquiera de las reivindicaciones 1 a 9.

25 13. Una célula recombinante, que comprende: una o más secuencias de ácido nucleico que codifican el NATNA de tipo cruzado con CRISPR de Clase 2 modificado de una cualquiera de las reivindicaciones 1 a 9.

30 14. Un kit, que comprende: el NATNA de tipo cruzado con CRISPR de Clase 2 modificado de una cualquiera de las reivindicaciones 1 a 9; y un tampón; en donde el kit comprende además preferentemente una proteína Cas9, una proteína Cpfl, o una proteína Cas9 y una proteína Cpfl.

35 15. Un método de unión al ADN, que comprende: poner en contacto una primera secuencia de ADN diana del ADN y una segunda secuencia de ADN diana del ADN con la composición de ácido nucleico /proteína de una cualquiera de las reivindicaciones 10 u 11, facilitando de esta manera la unión de la composición de ácido nucleico/proteína a la primera secuencia de ADN diana del ADN y la segunda secuencia de ADN diana del ADN, en el que el elemento espaciador de Cpfl-NATNA es complementario de la primera secuencia de ADN diana y el espaciador de Cas9-NATNA es complementario de la segunda secuencia de ADN diana, en donde el método no comprende un método de tratamiento del cuerpo humano o animal mediante terapia.

45 16. Un método de corte del ADN, que comprende: poner en contacto una primera secuencia de ADN diana del ADN y una segunda secuencia de ADN diana del ADN con la composición de ácido nucleico/proteína de la reivindicación 10, facilitando de esta manera la unión de la composición de ácido nucleico/proteína con la primera secuencia de ADN diana y la segunda secuencia de ADN diana; dando como resultado el corte de la primera secuencia de ADN diana y la segunda secuencia de ADN diana; en el que el elemento espaciador de Cpfl-NATNA es complementario de la primera secuencia de ADN diana y el espaciador de Cas9-NATNA es complementario de la segunda secuencia de ADN diana; y en el que la proteína Cas9 de la composición de ácido nucleico/proteína es capaz de cortar la segunda secuencia de ADN diana y la proteína Cpfl de la composición de ácido nucleico/proteína es capaz de cortar la primera secuencia de ADN diana, en donde el método no comprende un proceso para la modificación de la identidad genética de la línea germinal de un ser humano y, en donde el método no comprende un método de tratamiento del cuerpo humano o animal mediante terapia.

60

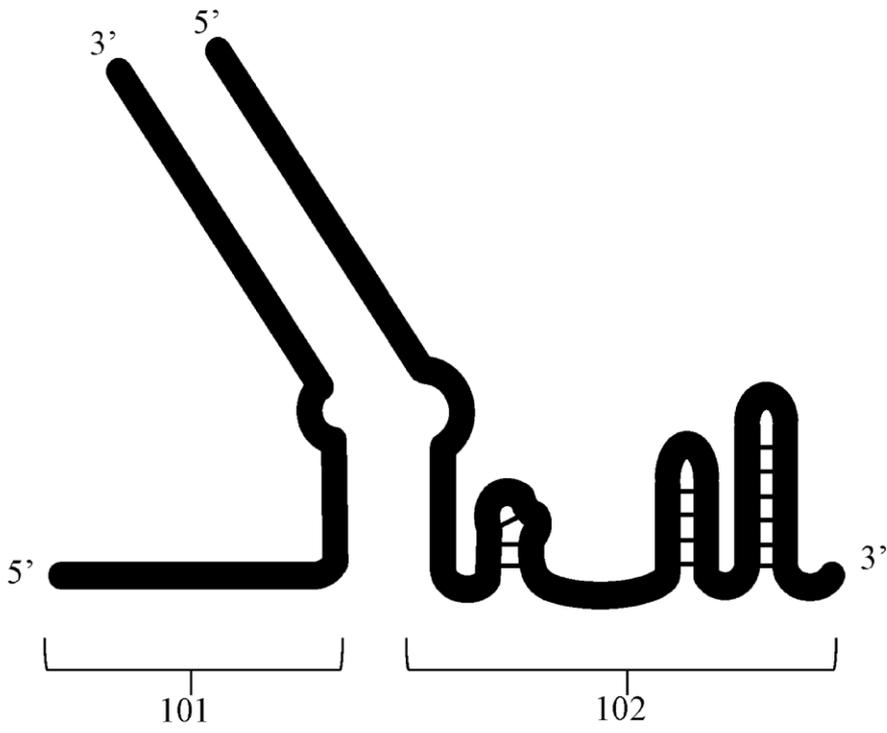


FIG. 1A

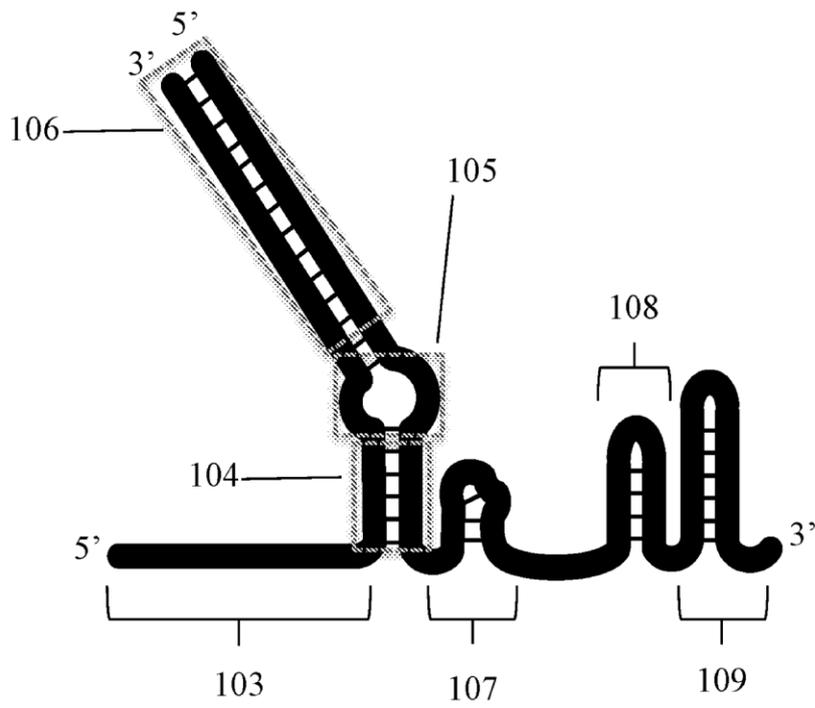


FIG. 1B

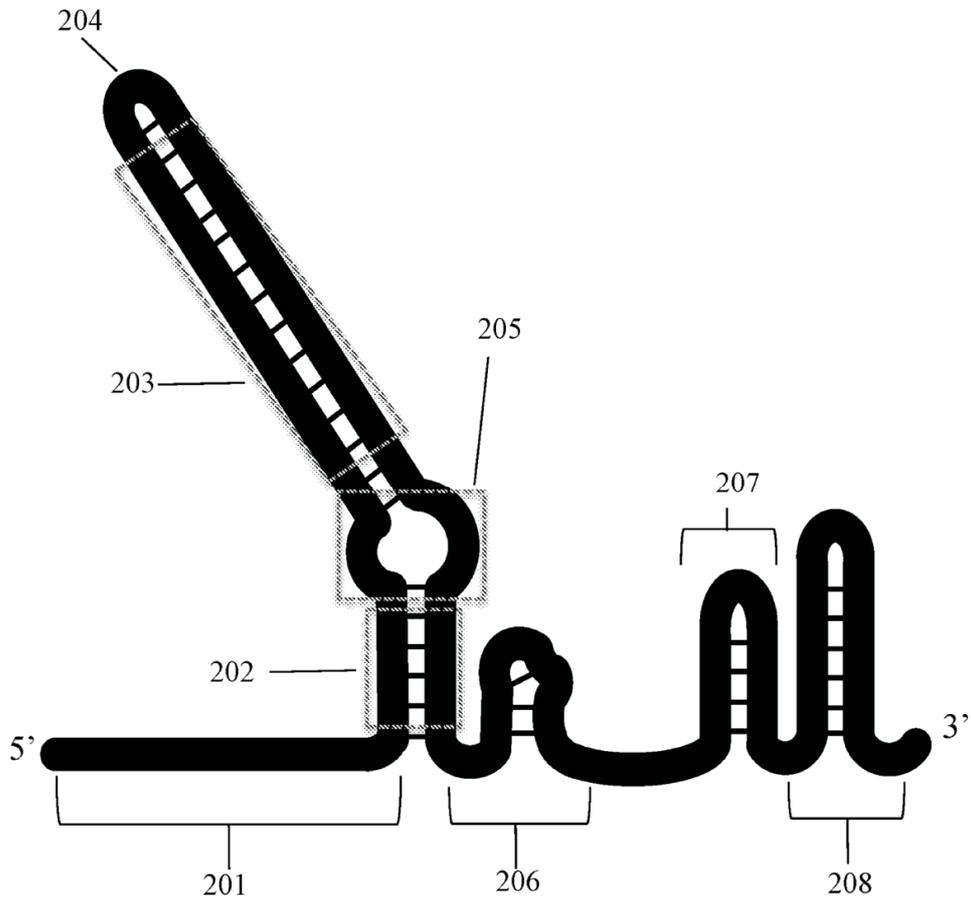


FIG. 2

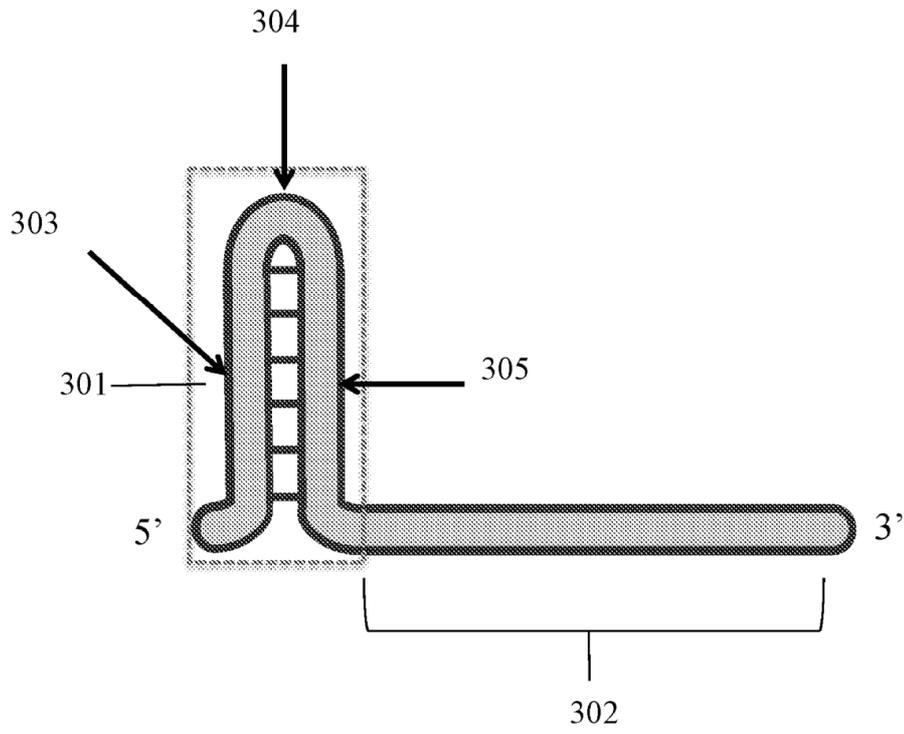


FIG. 3

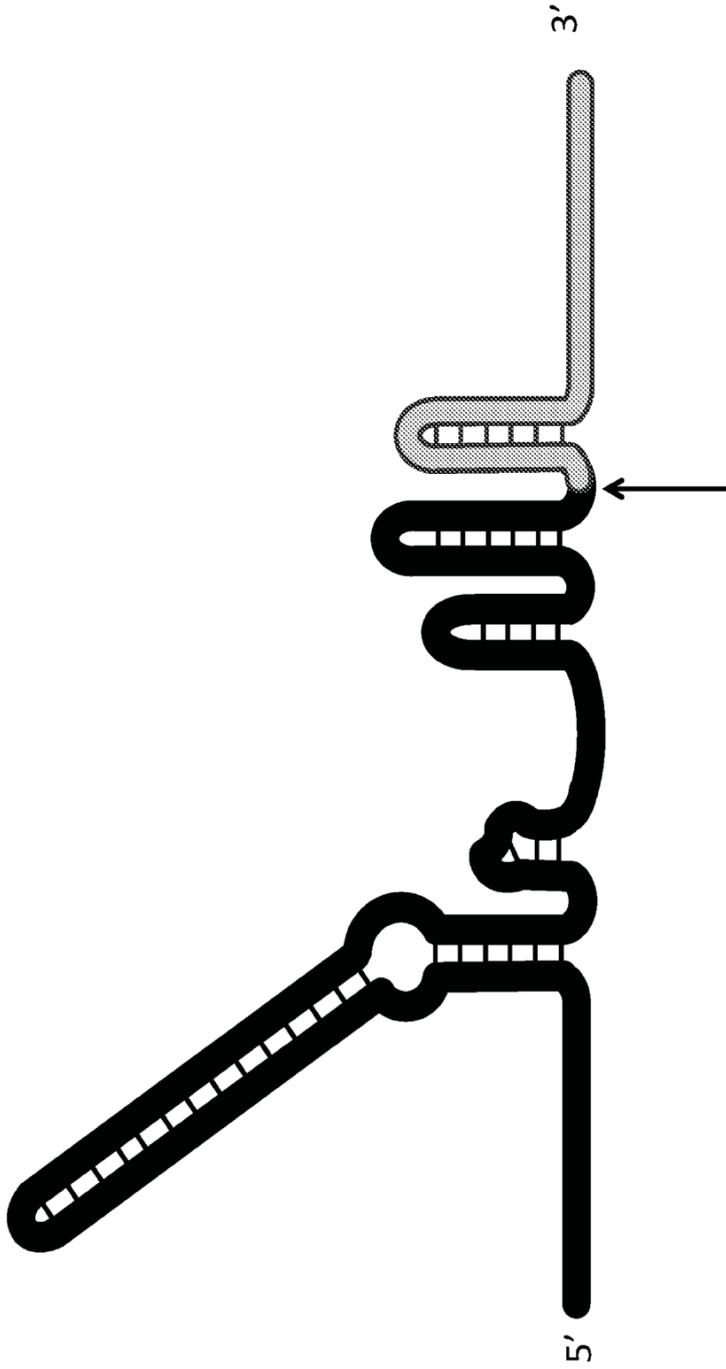


FIG. 4

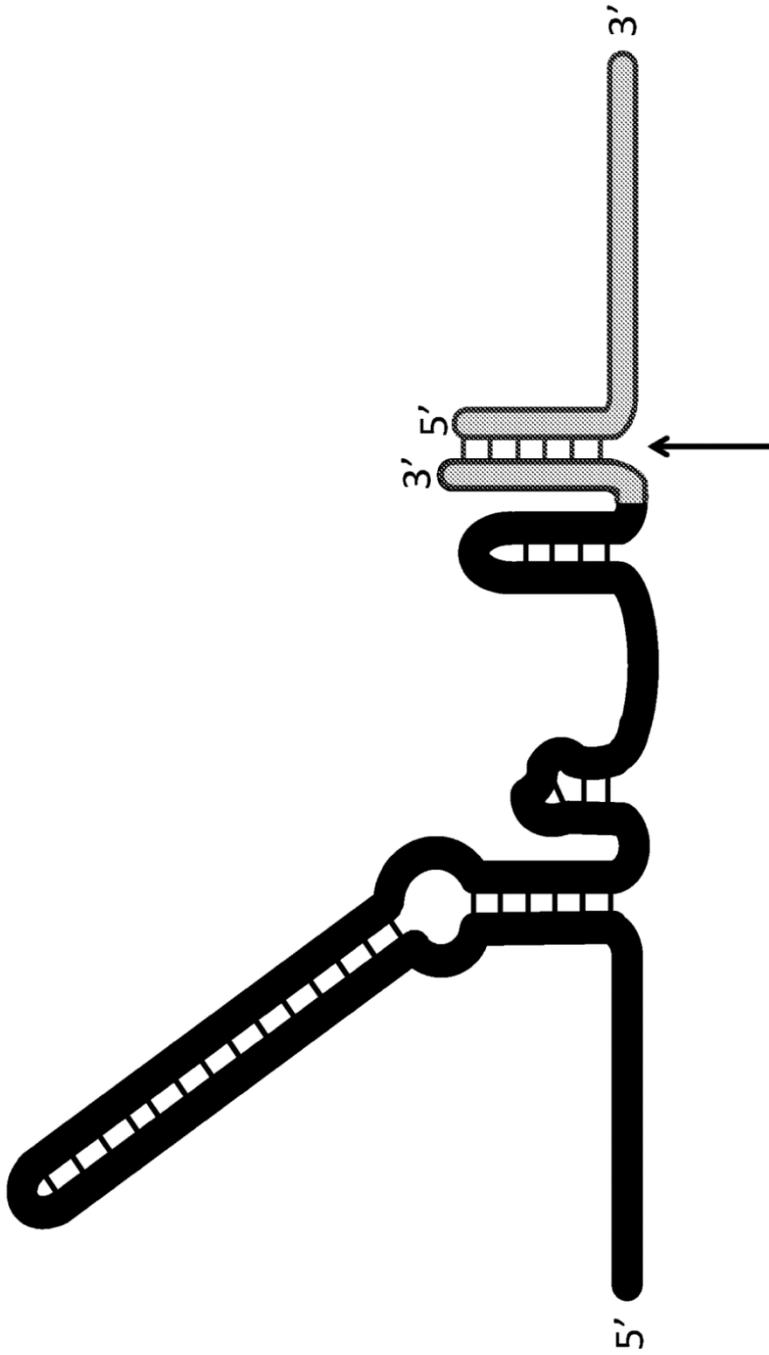


FIG. 5A

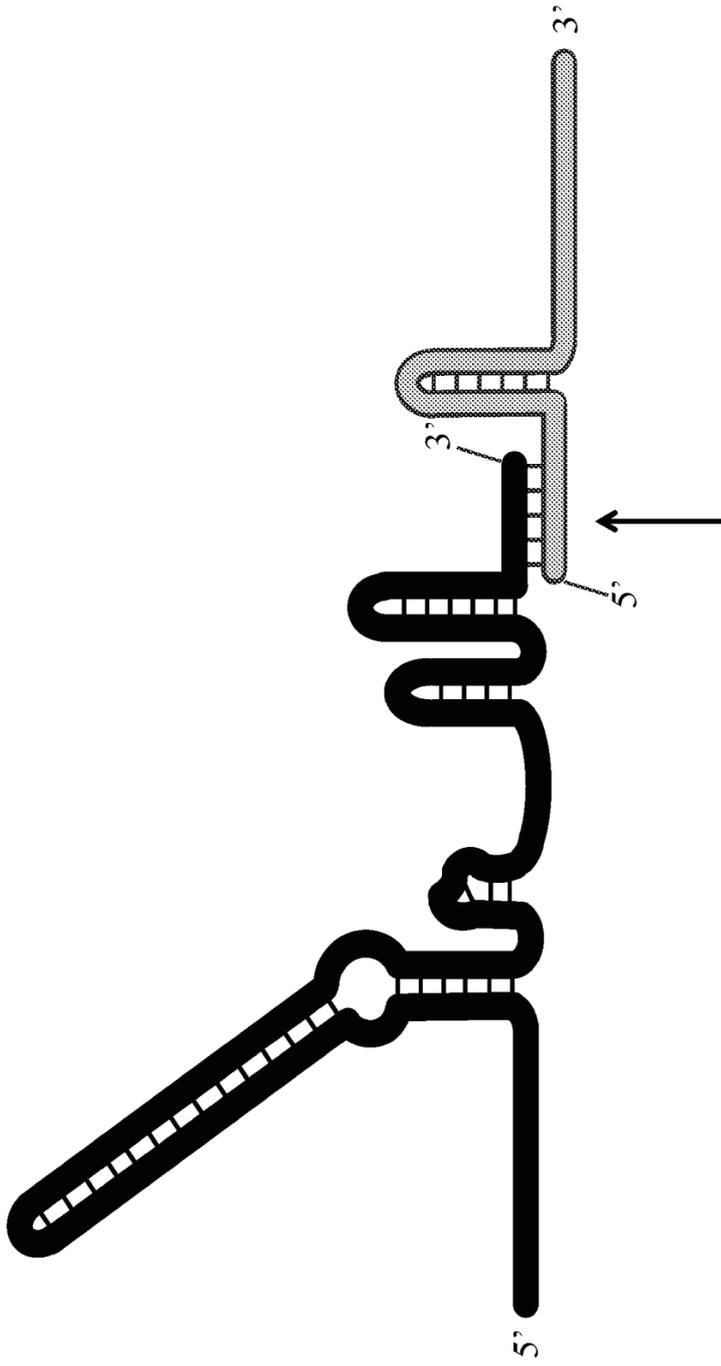


FIG. 5B

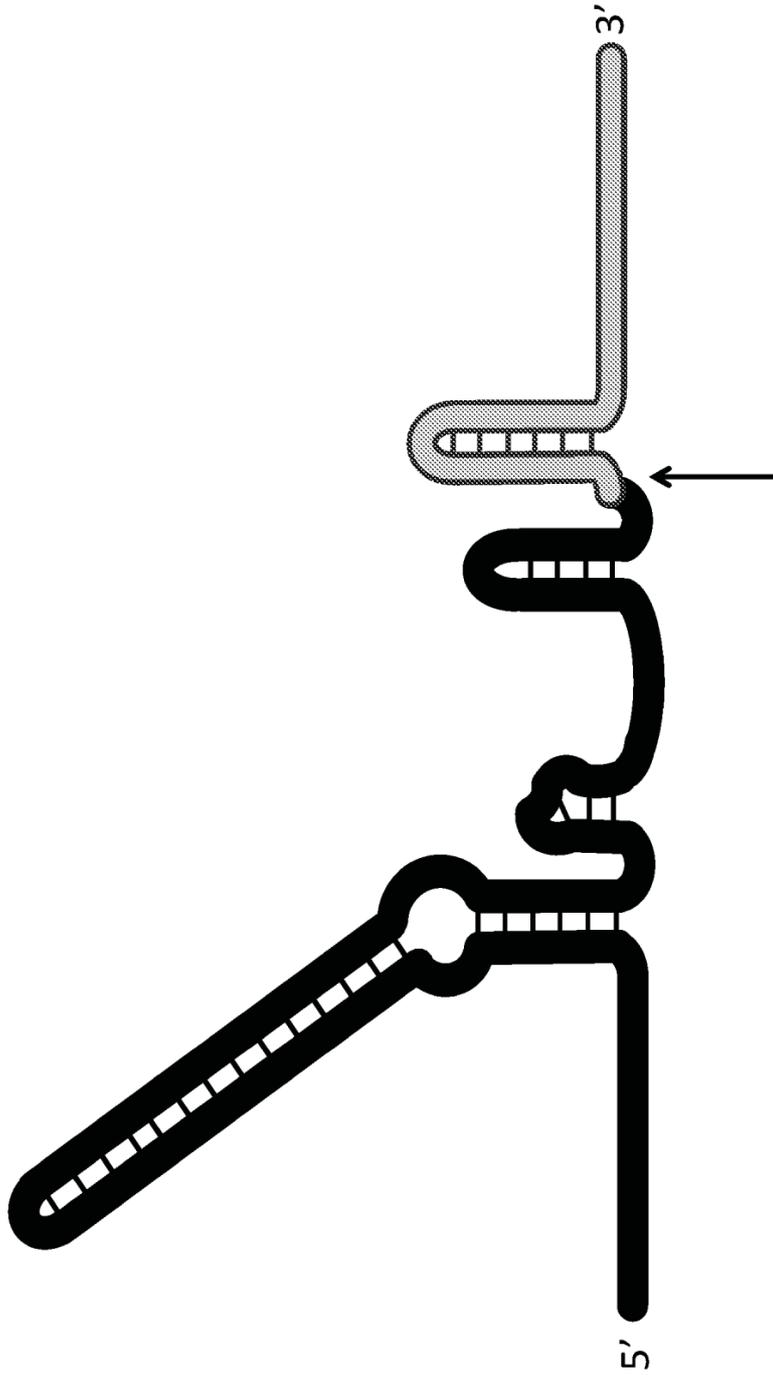


FIG. 6

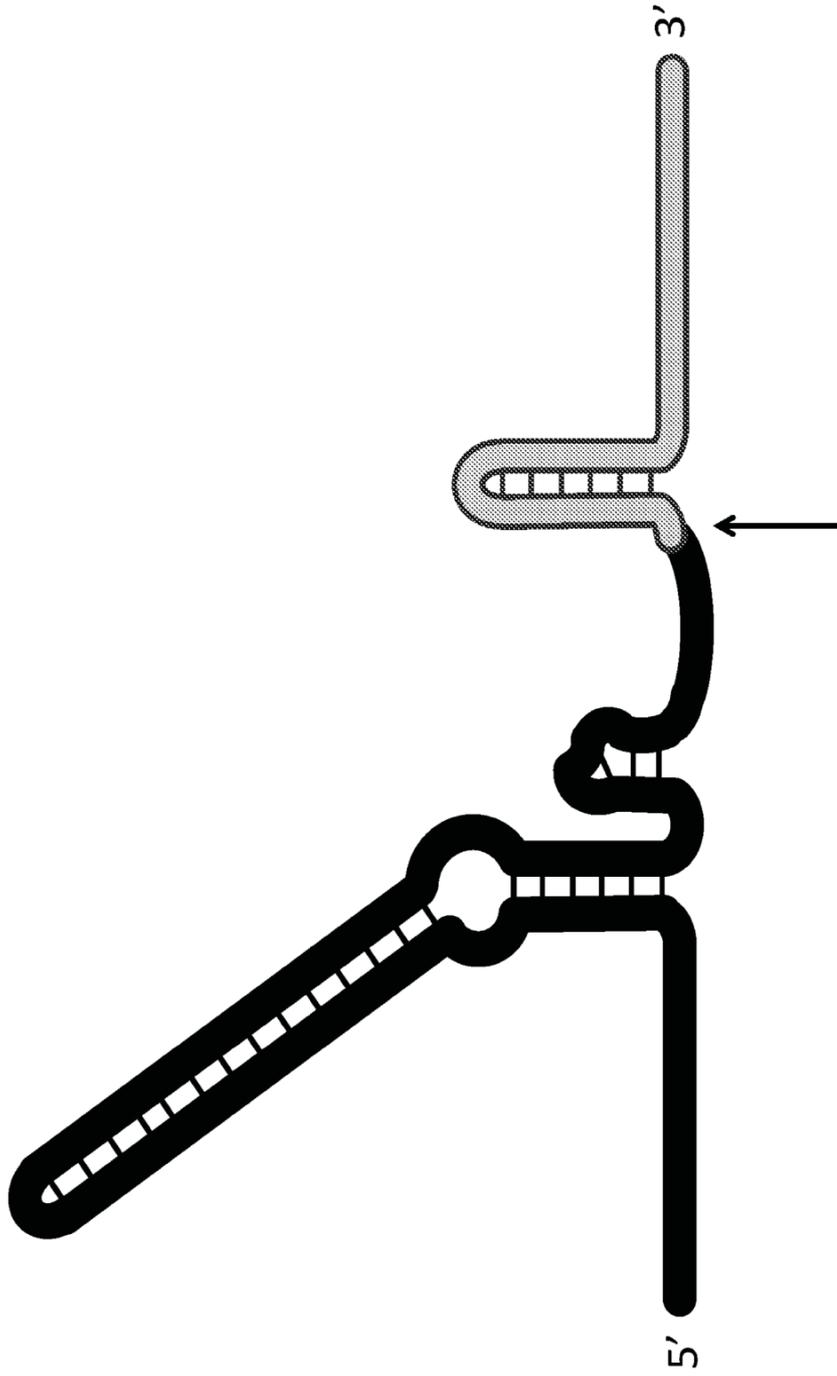


FIG. 7

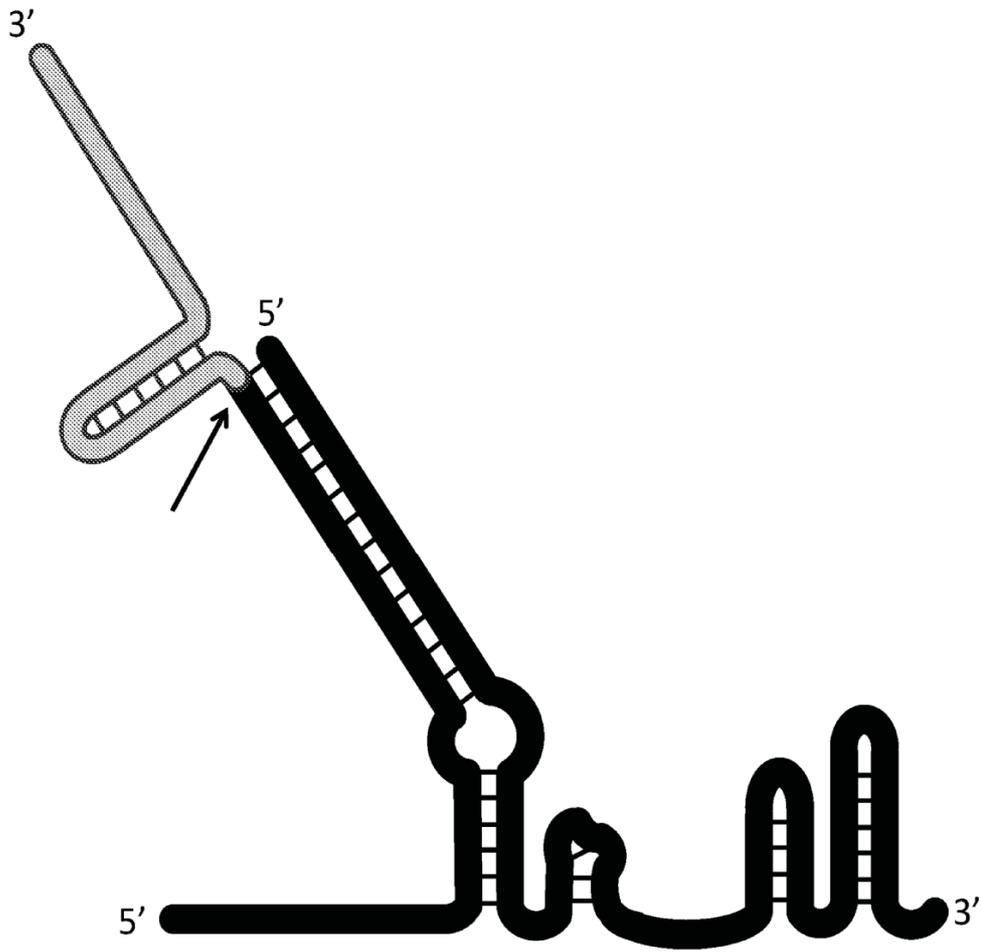


FIG. 8

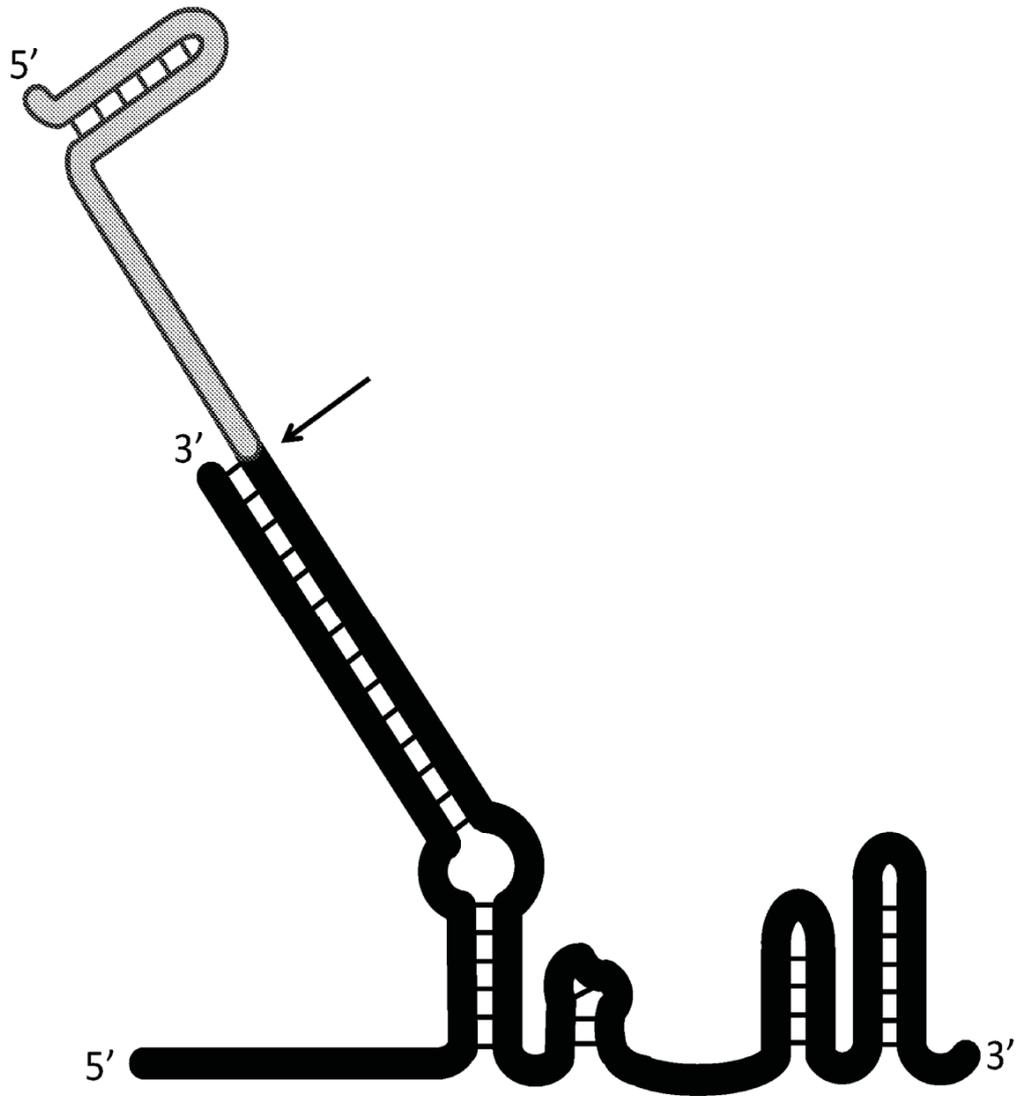


FIG. 9

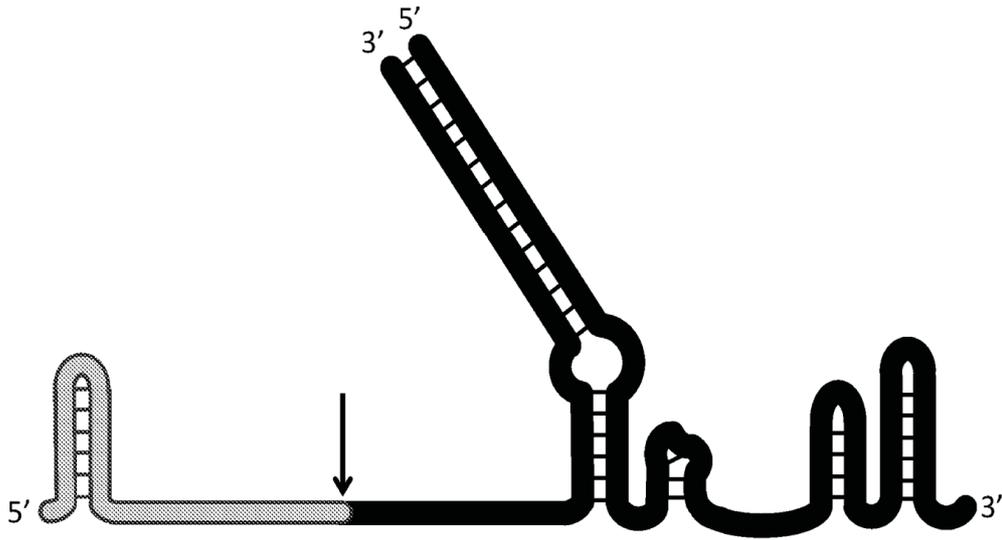


FIG. 10

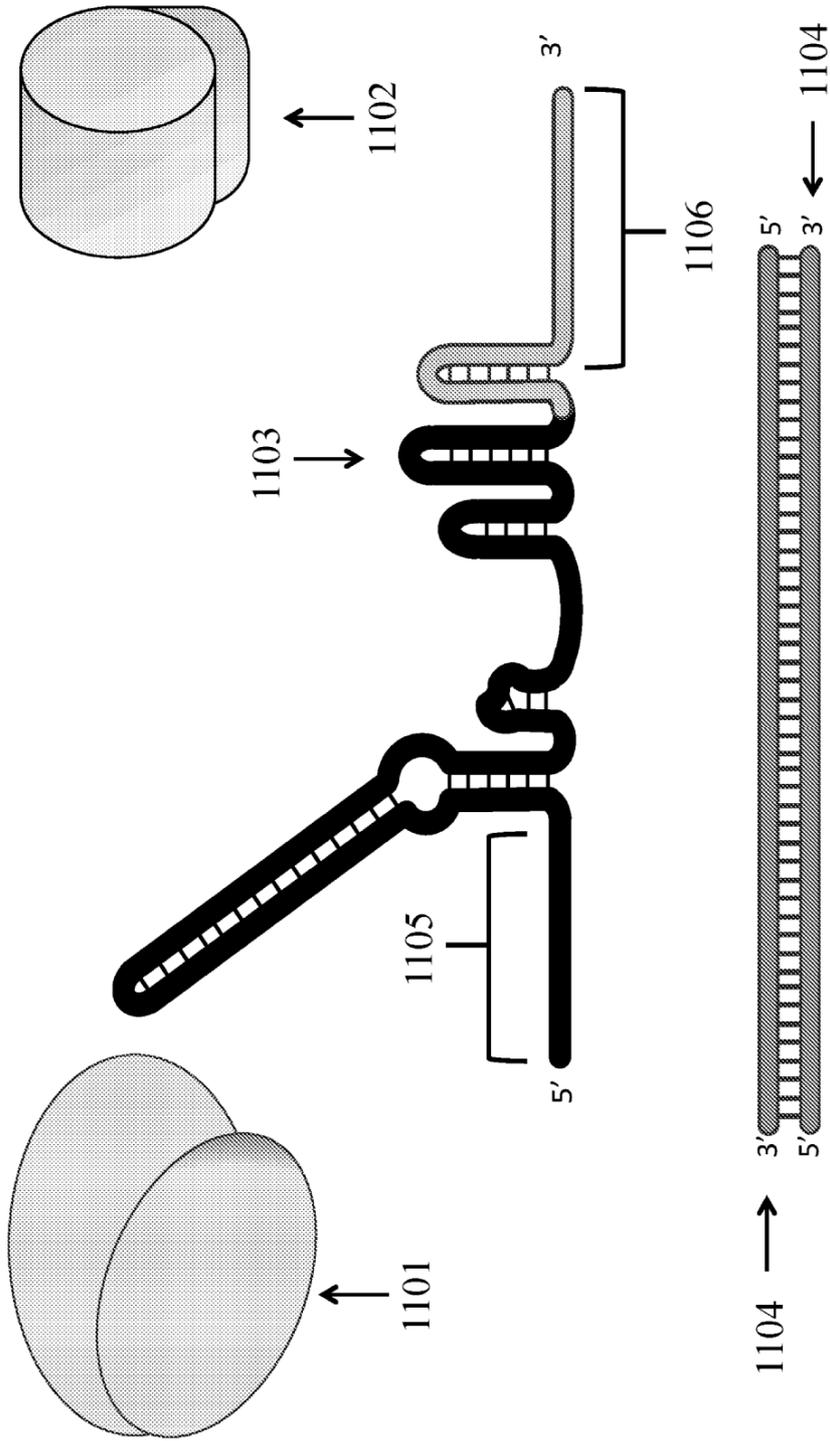


FIG. 11A



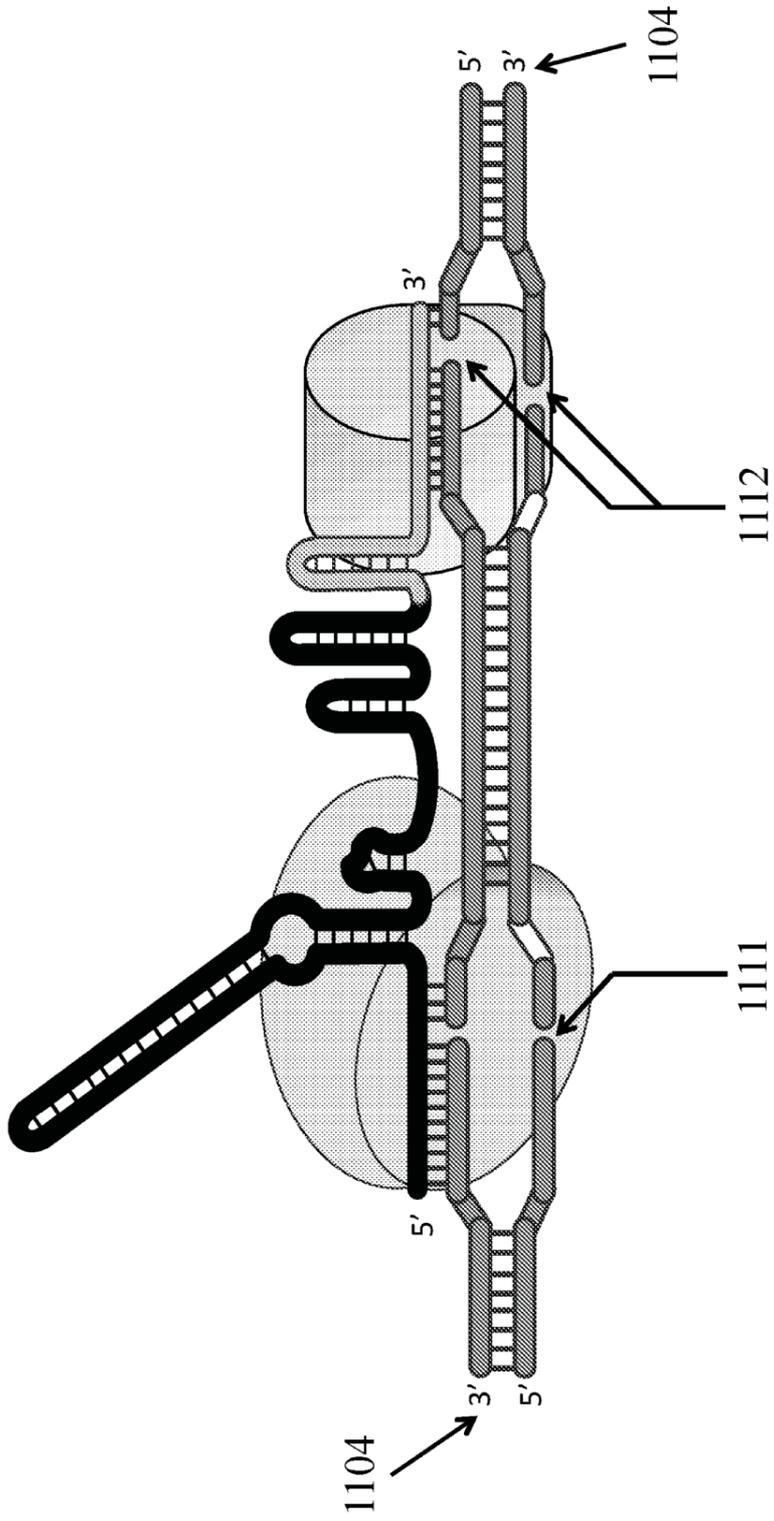


FIG. 11C

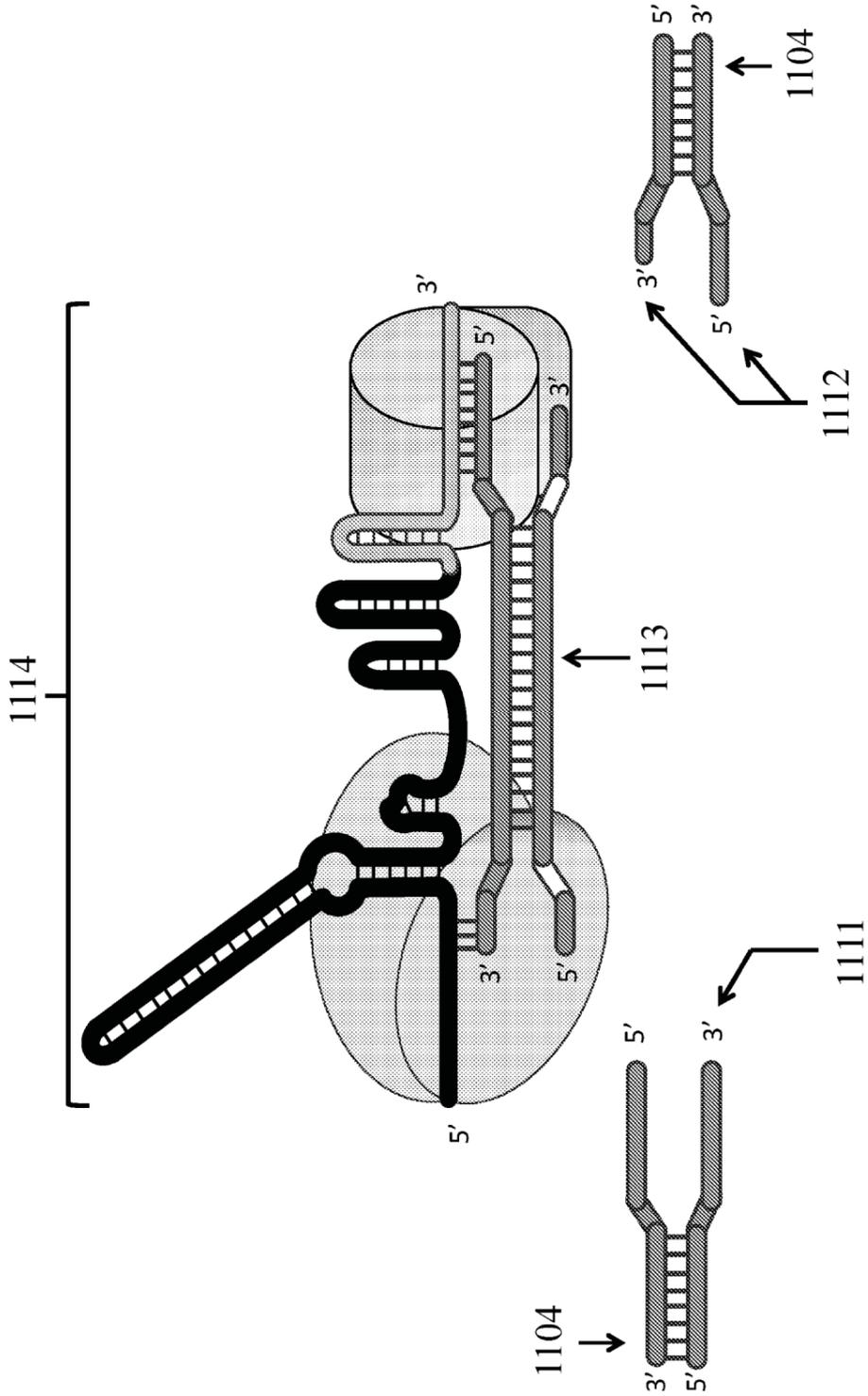


FIG. 11D

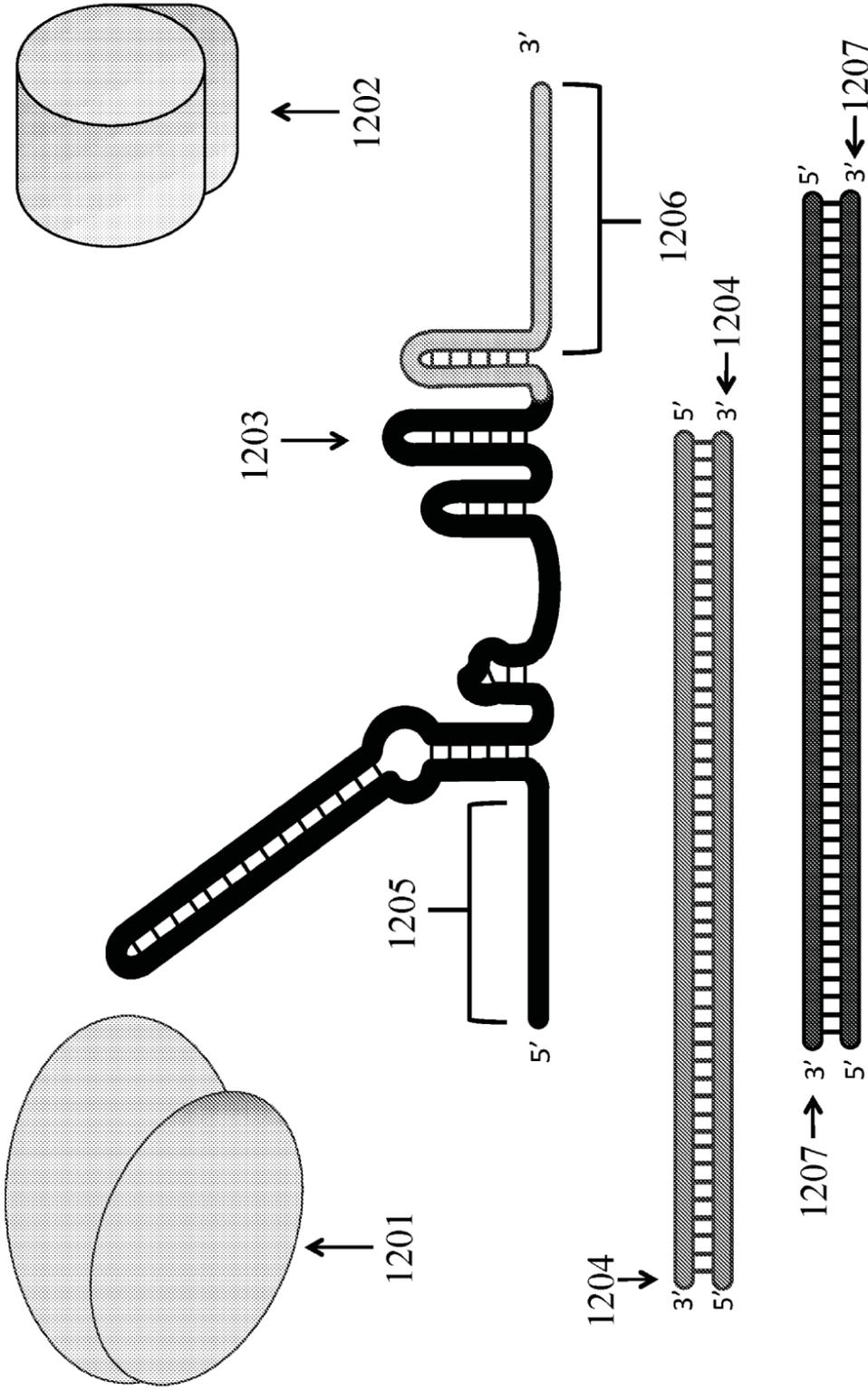


FIG. 12A

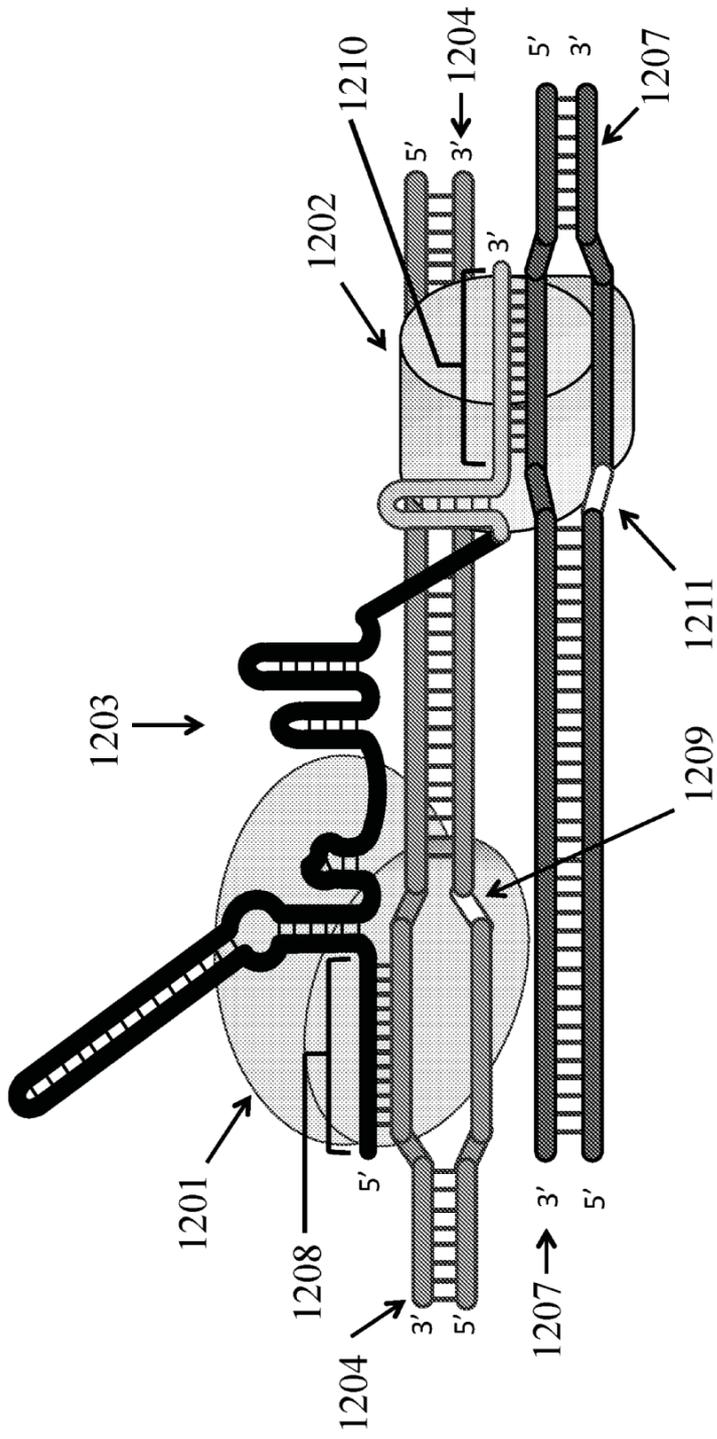


FIG. 12B

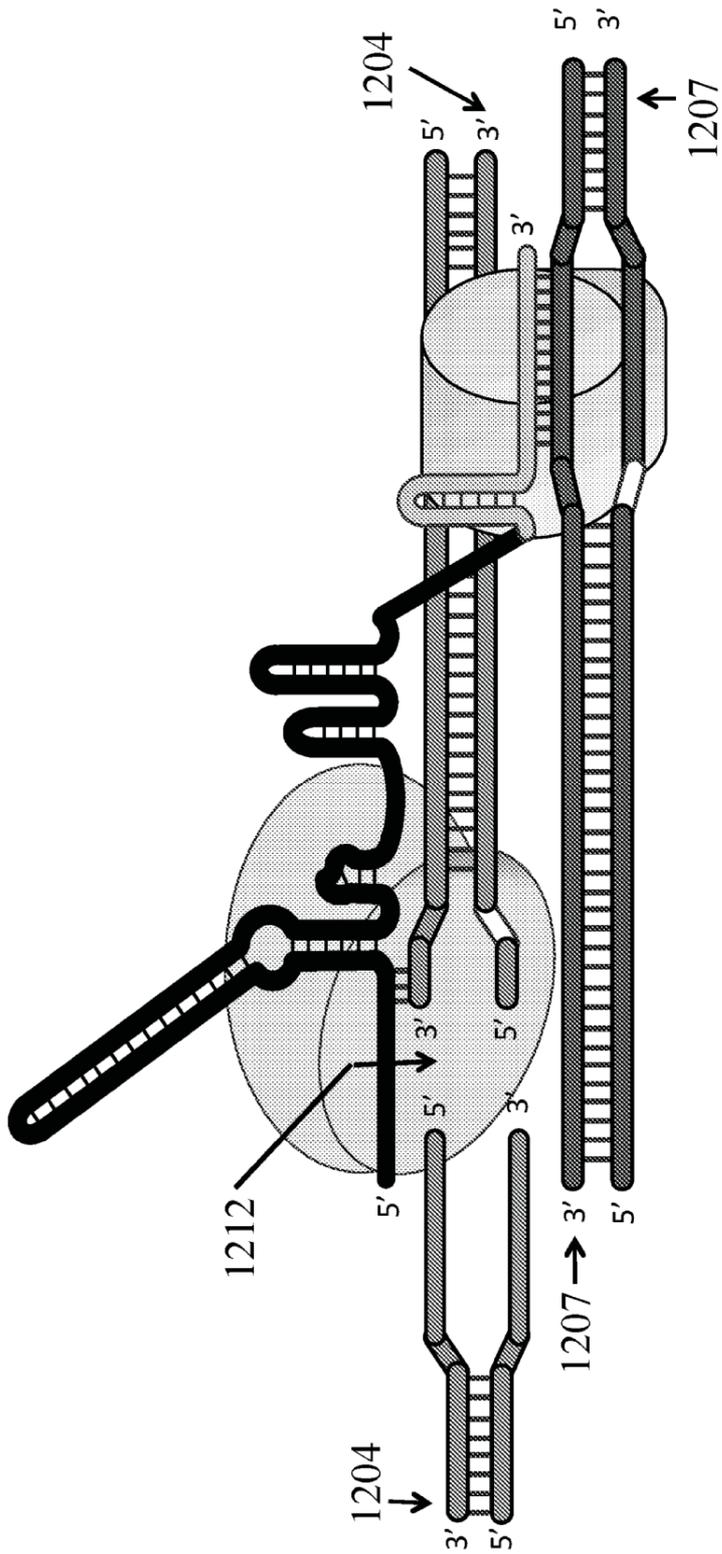


FIG. 12C

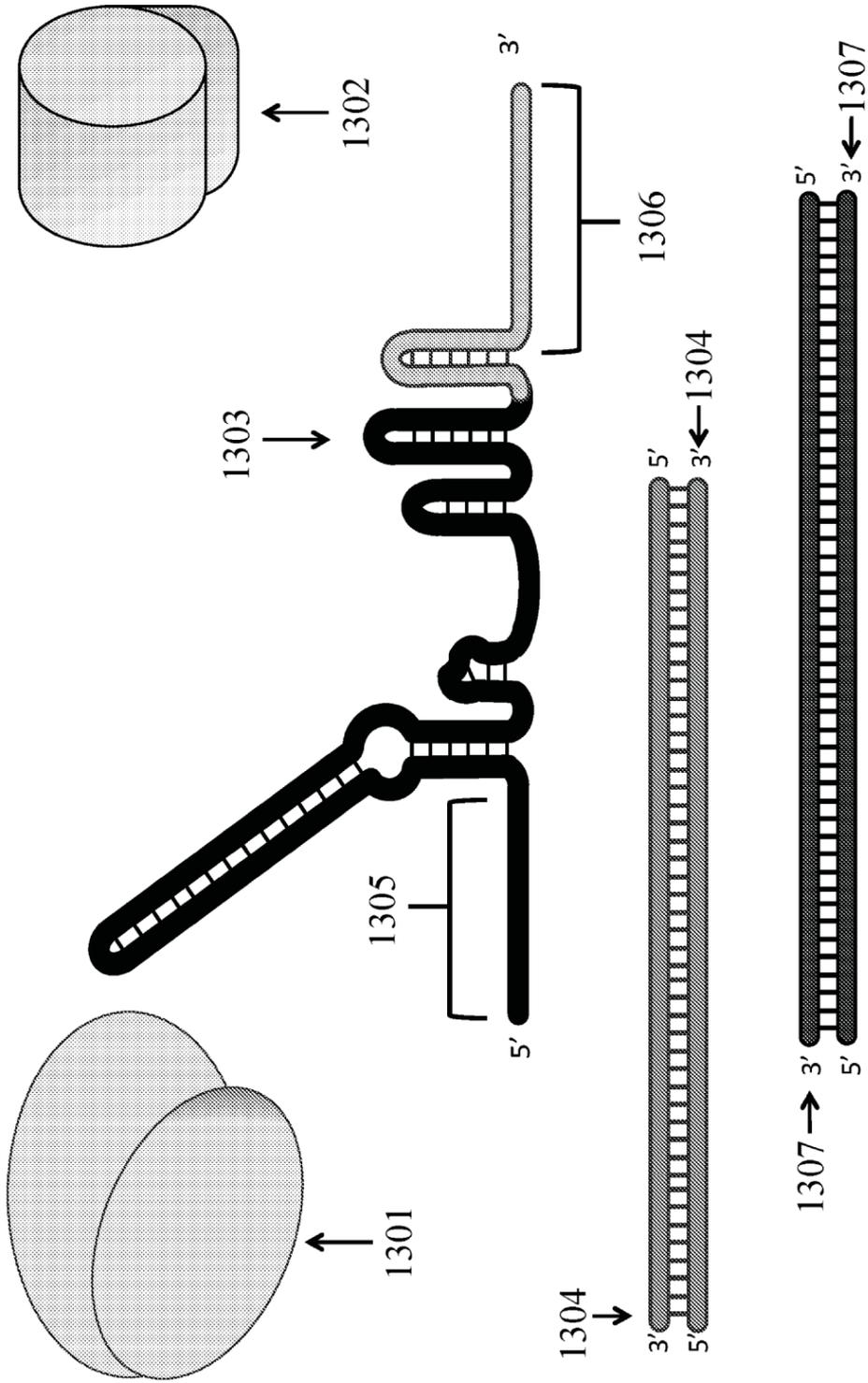


FIG. 13A

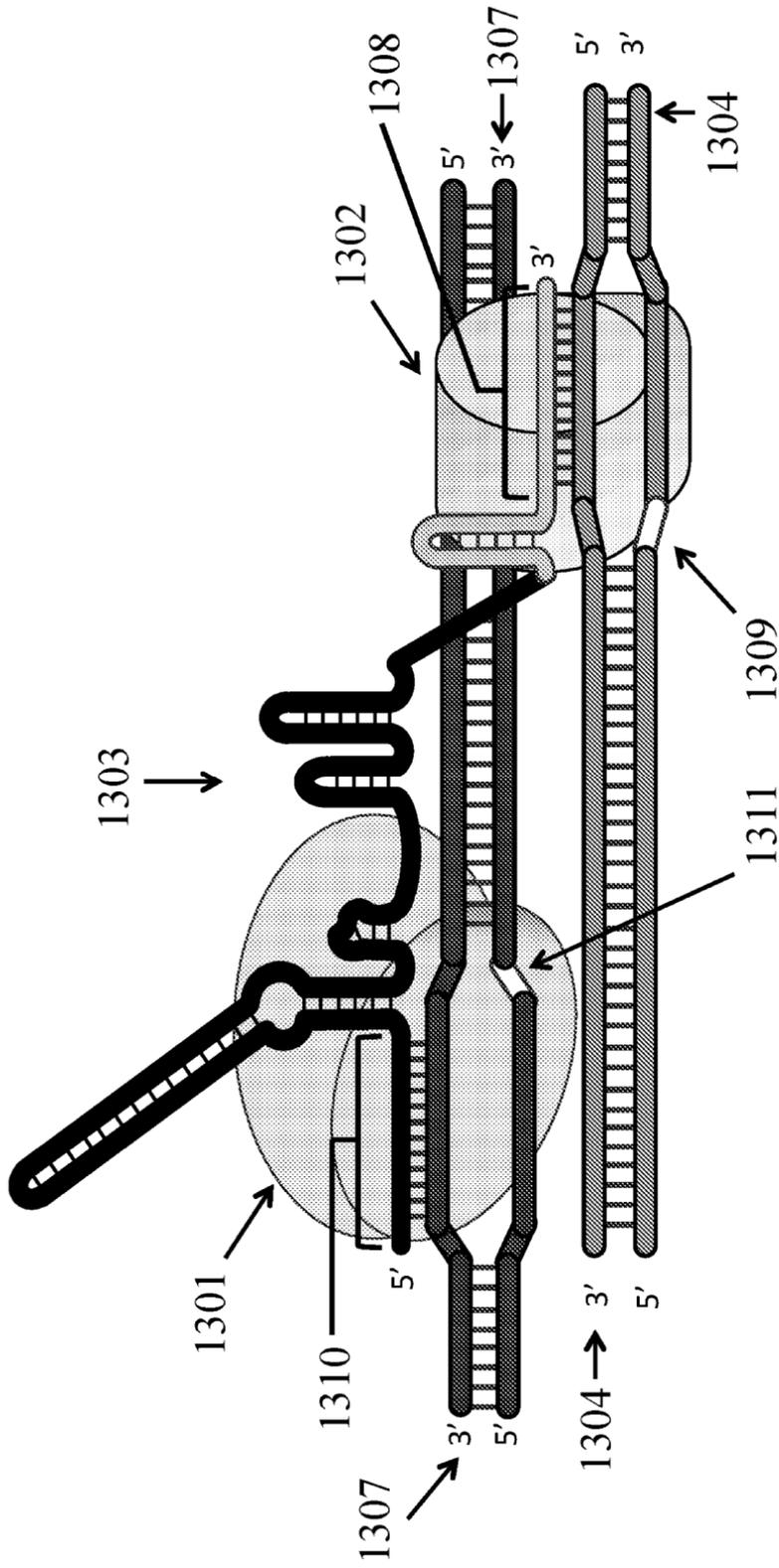


FIG. 13B

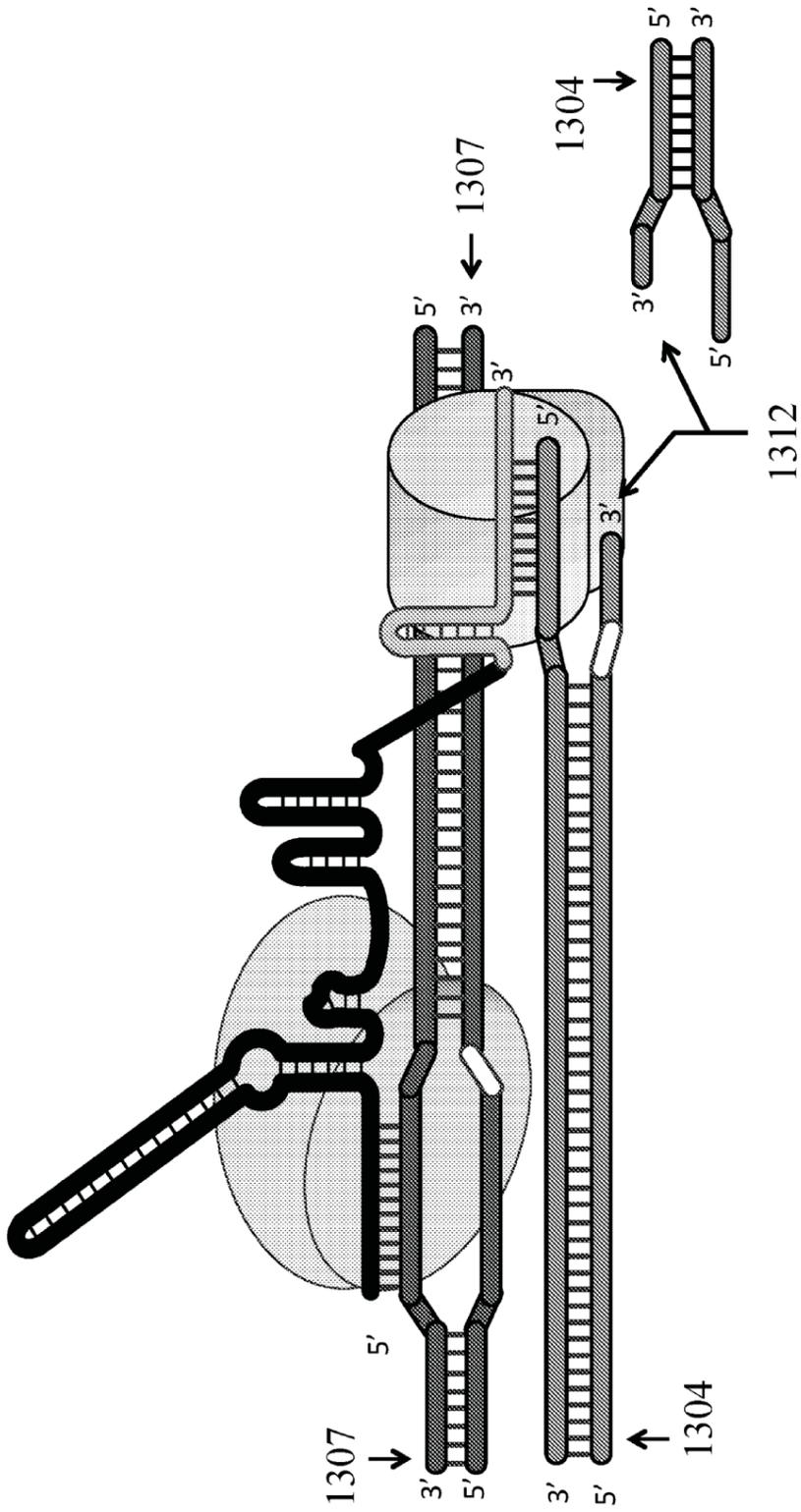


FIG. 13C



FIG. 14A

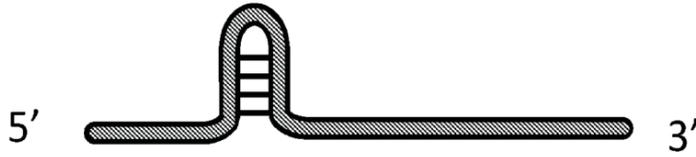


FIG. 14B

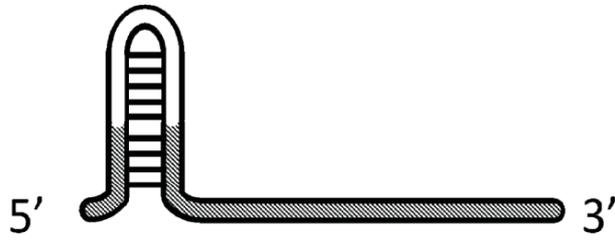


FIG. 14C

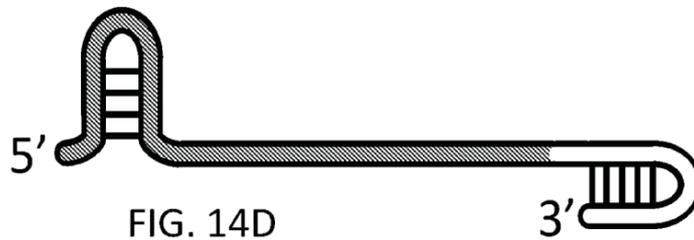


FIG. 14D

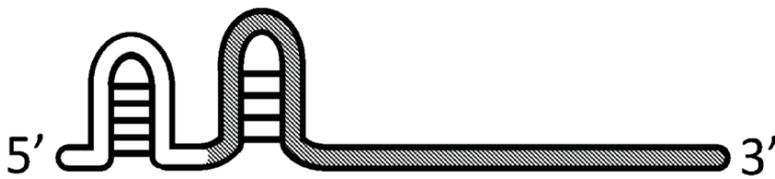


FIG. 14E

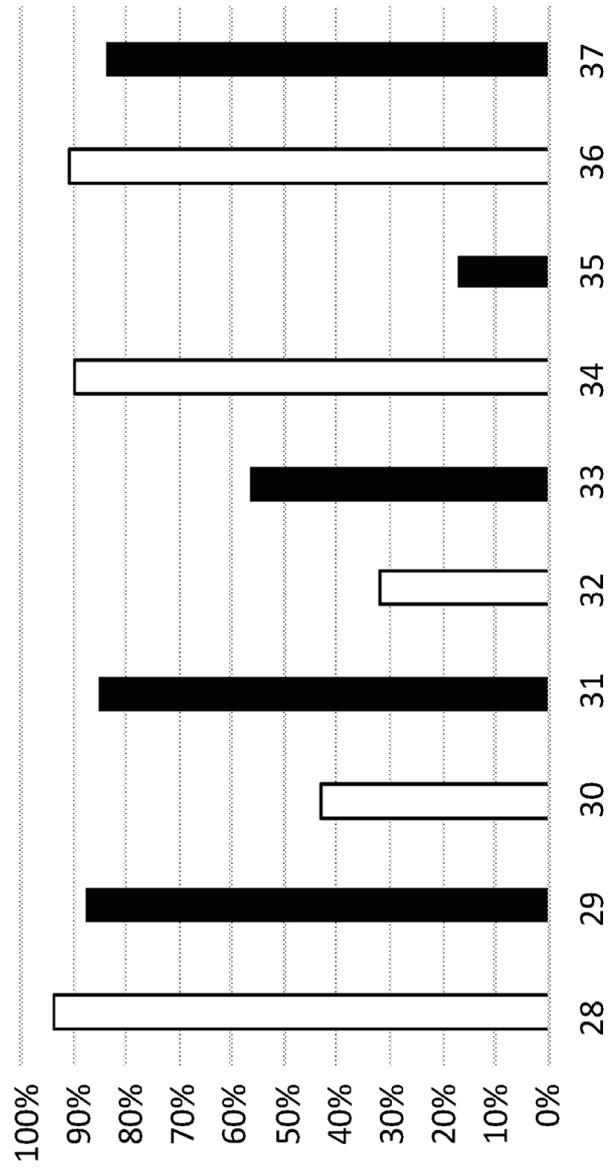


FIG. 15