(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 749 624**

(51) Int. CI.:

C12Q 1/6818 (2008.01)
C12Q 1/6869 (2008.01)
C12Q 1/6806 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **12.03.2014 PCT/US2014/023828**

(87) Fecha y número de publicación internacional: **25.09.2014 WO14150624**

(96) Fecha de presentación y número de la solicitud europea: **12.03.2014 E 14771135 (2)**

(97) Fecha y número de publicación de la concesión europea: **31.07.2019 EP 2971167**

(54) Título: **Composiciones y métodos de ácidos nucleicos dirigidos a ácido nucleico**

(30) Prioridad:

14.03.2013 US 201361781598 P
01.05.2013 US 201361818386 P
01.05.2013 US 201361818382 P
10.05.2013 US 201361822002 P
07.06.2013 US 201361832690 P
12.07.2013 US 201361845714 P
26.07.2013 US 201361858767 P
29.07.2013 US 201361859661 P
14.08.2013 US 201361865743 P
27.09.2013 US 201361883804 P
04.11.2013 US 201361899712 P
05.11.2013 US 201361900311 P
11.11.2013 US 201361902723 P
12.11.2013 US 201361903232 P
19.11.2013 US 201361906211 P
19.11.2013 US 201361906335 P
21.11.2013 US 201361907216 P
22.11.2013 US 201361907777 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
23.03.2020

(73) Titular/es:

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

(72) Inventor/es:

MAY, ANDREW, PAUL;
HAURWITZ, RACHEL, E.;
DOUDNA, JENNIFER, A.;
BERGER, JAMES, M.;
CARTER, MATTHEW, MERRILL y
DONOHOUE, PAUL

(74) Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

Observaciones:

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

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

Composiciones y métodos de ácidos nucleicos dirigidos a ácido nucleico

Antecedentes de la invención

El diseño del genoma puede referirse a la modificación del genoma mediante la eliminación, la inserción, la mutación o la sustitución de secuencias específicas de ácido nucleico. La modificación puede ser específica del gen o de la ubicación. El diseño del genoma puede usar nucleasas para cortar un ácido nucleico generando así un sitio para la modificación. También se contempla el diseño del ácido nucleico no genómico. Una proteína que contiene un dominio nucleasa puede unirse y escindir un ácido nucleico diana formando un complejo con un ácido nucleico dirigido a ácido nucleico. En un ejemplo, la escisión puede introducir roturas bicatenarias en el ácido nucleico diana. Un ácido nucleico puede ser reparado, por ejemplo, por la maquinaria endógena de unión de extremos no homólogos (NHEJ). En un ejemplo adicional, se puede insertar una pieza de ácido nucleico. Las modificaciones de ácidos nucleicos dirigidos a ácido nucleico y polipéptidos dirigidos a un sitio pueden introducir nuevas funciones para su uso en el diseño del genoma.

Jinek *et al.* (Science, 2012, 337(6096):816-812), Cong *et al.* (*Science*, 2013, 339(6121):819-823), Mali *et al.* (*Science*, 2013, 339(6121):823-826), El documento WO 2013/176772 A1 y el documento WO 2014/093712 A1 describen ácidos nucleicos dirigidos a ácido nucleico guía sencillo que comprenden un dúplex con una región protuberante seguida de un primer tallo y capaz de formar un complejo con una proteína Cas9 CRISPR de Tipo II. Dicha estructura dúplex conserva las interacciones de emparejamiento de bases que se producen entre el dúplex de ARNcrtra:ARNcr de *S. pyogenes*.

Sumario de la invención

La invención se define por las reivindicaciones adjuntas.

En un aspecto, la divulgación proporciona un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P del ácido nucleico dirigido a ácido nucleico. En algunos aspectos, el dominio P comienza cadena abajo de un último nucleótido emparejado de un dúplex entre una repetición CRISPR y una secuencia de ARNcrtra del ácido nucleico dirigido a ácido nucleico. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado comprende además una secuencia enlazadora. En algunos aspectos, la secuencia enlazadora une la repetición CRISPR y la secuencia de ARNcrtra. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado aislado. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado recombinante. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado está adaptado a hibridarse con un ácido nucleico diana. En algunos aspectos, el dominio P comprende 2 nucleótidos adyacentes. En algunos aspectos, el dominio P comprende 3 nucleótidos adyacentes. En algunos aspectos, el dominio P comprende 4 nucleótidos adyacentes. En algunos aspectos, el dominio P comprende 5 nucleótidos adyacentes. En algunos aspectos, el dominio P comprende 6 o más nucleótidos adyacentes. En algunos aspectos, el dominio P comienza 1 nucleótido cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, el dominio P comienza 2 nucleótidos cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, el dominio P comienza 3 nucleótidos cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, el dominio P comienza 4 nucleótidos cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, el dominio P comienza 5 nucleótidos cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, el dominio P comienza 6 o más nucleótidos cadena abajo del último nucleótido emparejado del dúplex. En algunos aspectos, la mutación comprende una o más mutaciones. En algunos aspectos, la una o más mutaciones son adyacentes entre sí. En algunos aspectos, la una o más mutaciones están separadas entre sí. En algunos aspectos, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se hibride con un motivo adyacente al protoespaciador diferente. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende al menos 4 nucleótidos. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende al menos 5 nucleótidos. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende al menos 6 nucleótidos. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende al menos 7 o más nucleótidos. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende dos regiones no adyacentes. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende tres regiones no adyacentes. En algunos aspectos, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un ácido nucleico diana con una constante de disociación menor que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunos aspectos, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un ácido nucleico diana con mayor especificidad que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunos aspectos, la mutación está adaptada para reducir la unión del ácido nucleico dirigido a ácido nucleico diseñado a una secuencia no específica en un ácido nucleico diana en comparación con un ácido nucleico dirigido a ácido nucleico no diseñado. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado comprende además dos horquillas, en donde una de las dos horquillas comprende un dúplex entre un polinucleótido que comprende al menos un 50 % de identidad con un ARN de CRISPR sobre 6 nucleótidos contiguos, y un polinucleótido que comprende al menos un 50 % de identidad con un ARNcrtra sobre 6 nucleótidos contiguos; y, en donde una de las dos horquillas está 3' de la

primera horquilla, en donde la segunda horquilla comprende un dominio P diseñado. En algunos aspectos, la segunda horquilla está adaptada para eliminar el dúplex cuando el ácido nucleico está en contacto con un ácido nucleico diana. En algunos aspectos, el dominio P está adaptado a: hibridarse con un primer polinucleótido, en donde el primer polinucleótido comprende una región del ácido nucleico dirigido a ácido nucleico diseñado, hibridarse con un segundo polinucleótido, en donde el segundo polinucleótido comprende un ácido nucleico diana y se hibrida específicamente con el primer o segundo polinucleótido. En algunos aspectos, el primer polinucleótido comprende al menos un 50 % de identidad con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, el primer polinucleótido está ubicado cadena abajo de un dúplex entre un polinucleótido que comprende al menos un 50 % de identidad con una repetición CRISPR sobre 6 nucleótidos contiguos, y un polinucleótido que comprende al menos un 50 % de identidad con una secuencia de ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, el segundo polinucleótido comprende un motivo adyacente al protoespaciador. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado está adaptado a unirse a un polipéptido dirigido a un sitio. En algunos aspectos, la mutación comprende una inserción de uno o más nucleótidos en el dominio P. En algunos aspectos, la mutación comprende la eliminación de uno o más nucleótidos del dominio P. En algunos aspectos, la mutación comprende la mutación de uno o más nucleótidos. En algunos aspectos, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico se hibride con un motivo adyacente al protoespaciador diferente. En algunos aspectos, el motivo adyacente al protoespaciador diferente comprende un motivo adyacente al protoespaciador seleccionado del grupo que consiste en: 5'-NGGNG-3', 5'-NNAAAAW-3', 5'-NNNNGATT-3', 5'-GNNNCNNA-3' y 5'-NNNACA-3', o cualquier combinación de los mismos. En algunos aspectos, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una con una constante de disociación menor que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunos aspectos, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una con mayor especificidad que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunos aspectos, la mutación está configurada para reducir la unión del ácido nucleico dirigido a ácido nucleico diseñado a una secuencia no específica en un ácido nucleico diana en comparación con un ácido nucleico dirigido a ácido nucleico no diseñado.

En un aspecto, la divulgación proporciona un método de modificación de un ácido nucleico diana que comprende poner en contacto un ácido nucleico diana con un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P del ácido nucleico dirigido a ácido nucleico, y modificar el ácido nucleico diana. En algunos aspectos, el método comprende además insertar un polinucleótido donante en el ácido nucleico diana. En algunos aspectos, la modificación comprende escindir el ácido nucleico diana. En algunos aspectos, la modificación comprende modificar la transcripción del ácido nucleico diana.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P del ácido nucleico dirigido a ácido nucleico.

En un aspecto, la divulgación proporciona un kit que comprende: un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P del ácido nucleico dirigido a ácido nucleico; y un tampón. En algunos aspectos, el kit comprende además un polipéptido dirigido a un sitio. En algunos aspectos, el kit comprende además un polinucleótido donante. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico. En algunas realizaciones, la protuberancia se encuentra dentro de un dúplex entre una repetición CRISPR y una secuencia de ARNcrtra del ácido nucleico dirigido a ácido nucleico. En algunas realizaciones, el ácido nucleico dirigido a ácido nucleico diseñado comprende además una secuencia enlazadora. En algunas realizaciones, la secuencia enlazadora une la repetición CRISPR y la secuencia de ARNcrtra. En algunas realizaciones, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado aislado. En algunas realizaciones, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado recombinante. En algunas realizaciones, la protuberancia comprende al menos un 1 nucleótido no emparejado en la repetición CRISPR, y 1 nucleótido no emparejado en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos un 1 nucleótido no emparejado en la repetición CRISPR, y al menos 2 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos un 1 nucleótido no emparejado en la repetición CRISPR, y al menos 3 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos un 1 nucleótido no emparejado en la repetición CRISPR, y al menos 4 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos un 1 nucleótido no emparejado en la repetición CRISPR, y al menos 5 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos 2 nucleótidos no emparejados en la repetición CRISPR, y 1 nucleótido no emparejado en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos 3 nucleótidos no emparejados en la repetición CRISPR, y al menos 2 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos 4 nucleótidos no emparejados en la repetición CRISPR, y al menos 3 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos 5 nucleótidos no emparejados en la repetición CRISPR, y al menos 4 nucleótidos no emparejados en la secuencia de ARNcrtra. En algunas realizaciones, la protuberancia comprende al menos un nucleótido en la repetición CRISPR adaptado a formar

un par oscilante con al menos un nucleótido en la secuencia de ARNcrtra. En algunas realizaciones, la mutación comprende una o más mutaciones. En algunas realizaciones, la una o más mutaciones son adyacentes entre sí. En algunas realizaciones, la una o más mutaciones están separadas entre sí. En algunas realizaciones, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un polipéptido dirigido a un sitio diferente. En algunas realizaciones, el polipéptido dirigido a un sitio diferente es un homólogo de Cas9. En algunas realizaciones, el polipéptido dirigido a un sitio diferente es una versión mutada de Cas9. En algunas realizaciones, el polipéptido dirigido a un sitio diferente comprenden una identidad de secuencia de aminoácidos del 10 % con Cas9 en un dominio nucleasa seleccionado del grupo que consiste en: un dominio nucleasa RuvC y un dominio nucleasa HNH, o cualquier combinación de los mismos. En algunas realizaciones, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se hibride con un motivo adyacente al protoespaciador diferente. En algunas realizaciones, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un polipéptido dirigido a un sitio con una constante de disociación menor que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un polipéptido dirigido a un sitio con mayor especificidad que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está adaptada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado dirija a un polipéptido dirigido a un sitio para escindir un ácido nucleico diana con mayor especificidad que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está adaptada para reducir la unión del ácido nucleico dirigido a ácido nucleico diseñado a una secuencia no específica en un ácido nucleico diana en comparación con un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, el ácido nucleico dirigido a ácido nucleico diseñado está adaptado a hibridarse con un ácido nucleico diana. En algunas realizaciones, la mutación comprende la inserción de uno o más nucleótidos en la protuberancia. En algunas realizaciones, la mutación comprende la eliminación de uno o más nucleótidos de la protuberancia. En algunas realizaciones, la mutación comprende la mutación de uno o más nucleótidos. En algunas realizaciones, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se hibride con un motivo adyacente al protoespaciador diferente en comparación con un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un polipéptido dirigido a un sitio con una constante de disociación menor que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está configurada para permitir que el ácido nucleico dirigido a ácido nucleico diseñado se una a un polipéptido dirigido a un sitio con mayor especificidad que un ácido nucleico dirigido a ácido nucleico no diseñado. En algunas realizaciones, la mutación está configurada para reducir la unión del ácido nucleico dirigido a ácido nucleico diseñado a una secuencia no específica en un ácido nucleico diana en comparación con un ácido nucleico dirigido a ácido nucleico no diseñado.

En un aspecto, la divulgación proporciona un método de modificación de un ácido nucleico diana que comprende: poner en contacto el ácido nucleico diana con un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; y modificar el ácido nucleico diana. En algunos aspectos, el método comprende además insertar un polinucleótido donante en el ácido nucleico diana. En algunas realizaciones, la modificación comprende escindir el ácido nucleico diana. En algunos aspectos, la modificación comprende modificar la transcripción del ácido nucleico diana.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; y modificar el ácido nucleico diana.

En un aspecto, la divulgación proporciona un kit que comprende: un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; y modificar el ácido nucleico diana; y un tampón. En algunas realizaciones, el kit comprende además un polipéptido dirigido a un sitio. En algunas realizaciones, el kit comprende además un polinucleótido donante. En algunas realizaciones, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un método de producción de una célula marcada con polinucleótido donante que comprende: escindir un ácido nucleico diana de una célula usando un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, insertar un polinucleótido donante en un ácido nucleico diana escindido, propagar la célula portadora del polinucleótido donante y determinar un origen de la célula marcada con el polinucleótido donante. En algunos aspectos, el método se realiza *in vivo*. En algunas realizaciones, el método se realiza *in vitro*. En algunos aspectos, el método se realiza *in situ*. En algunas realizaciones, la propagación produce una población de células. En algunas realizaciones, la propagación produce una estirpe celular. En algunas realizaciones, el método comprende además determinar una secuencia de ácido nucleico de un ácido nucleico de la célula. En algunas realizaciones, la secuencia de ácido nucleico determina un origen de la célula. En algunas realizaciones, la determinación comprende determinar un genotipo de la célula. En algunas realizaciones, la propagación comprende diferenciar la célula. En algunas realizaciones, la propagación comprende desdiferenciar la célula. En algunas realizaciones, la propagación comprende diferenciar la célula y luego desdiferenciar la célula. En algunas realizaciones, la propagación comprende diferenciar la célula. En algunas realizaciones, la propagación comprende inducir a la célula a dividirse. En algunas realizaciones, la propagación comprende inducir a la célula a entrar en el ciclo celular. En algunas realizaciones, la propagación comprende que la célula forma una metástasis. En algunas realizaciones, la propagación comprende diferenciar una célula pluripotente en una célula diferenciada. En

algunas realizaciones, la célula es una célula diferenciada. En algunas realizaciones, la célula es una célula desdiferenciada. En algunas realizaciones, la célula es una célula madre. En algunas realizaciones, la célula es una célula madre pluripotente. En algunas realizaciones, la célula es una estirpe celular eucariota. En algunas realizaciones, la célula es una estirpe celular primaria. En algunas realizaciones, la célula es una estirpe celular derivada del paciente. En algunos aspectos, el método comprende además trasplantar la célula a un organismo. En algunos aspectos, el organismo es un ser humano. En algunos aspectos, el organismo es un mamífero. En algunos aspectos, el organismo se selecciona del grupo que consiste en: un ser humano, un perro, una rata, un ratón, un pollo, un pez, un gato, una planta y un primate. En algunas realizaciones, el método comprende además seleccionar la célula. En algunas realizaciones, el polinucleótido donante se inserta en un ácido nucleico diana que se expresa en un estado celular. En algunas realizaciones, el polinucleótido donante se inserta en un ácido nucleico diana que se expresa en una pluralidad de tipos de células. En algunas realizaciones, el polinucleótido del donante se inserta en un ácido nucleico diana que se expresa en un estado pluripotente. En algunas realizaciones, el polinucleótido donante se inserta en un ácido nucleico diana que se expresa en un estado diferenciado.

En un aspecto, la divulgación proporciona un método de creación de una estirpe celular expandida clónicamente que comprende: introducir en una célula un complejo que comprende: un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, poner en contacto el complejo con un ácido nucleico diana, escindir el ácido nucleico diana, en donde la escisión es realizada por el complejo, produciendo así un ácido nucleico diana escindido, insertar un polinucleótido donante en el ácido nucleico diana escindido, propagar la célula, en donde la propagación produce la estirpe celular expandida clónicamente. En algunos aspectos, la célula se selecciona del grupo que consiste en: célula HeLa, célula de ovario de hámster chino, célula 293-T, un feocromocitoma, un fibroblasto de neuroblastomas, un rabdomiosarcoma, una célula del ganglio de la raíz dorsal, una célula NSO, CV-I (ATCC CCL 70), COS-I (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C 1271 (ATCC CRL-1616), BS-C-I (ATCC CCL 26), MRC-5 (ATCC CCL 171), células L, HEK-293 (ATCC CRL1 573) y PC 12 (ATCC CRL-1721), HEK293T (ATCC CRL-11268), RBL (ATCC CRL-1378), SH-SY5Y (ATCC CRL-2266), MDCK (ATCC CCL-34), SJ-RH30 (ATCC CRL-2061), HepG2 (ATCC HB-8065), ND7/23 (ECACC 92090903), CHO (ECACC 85050302), Vera (ATCC CCL 81), Caco-2 (ATCC HTB 37), K562 (ATCC CCL 243), Jurkat (ATCC TIB-152), Per.Có, Huvec (ATCC PCS 100-010 primaria humana, CRL 2514 de ratón, CRL 2515, CRL 2516), HuH-7D12 (ECACC 01042712), 293 (ATCC CRL 10852), A549 (ATCC CCL 185), IMR-90 (ATCC CCL 186), MCF-7 (ATC HTB-22), U-2 OS (ATCC HTB-96) y T84 (ATCC CCL 248), o cualquier combinación de las mismas. En algunos aspectos, la célula es una célula madre. En algunos aspectos, la célula es una célula diferenciada. En algunos aspectos, la célula es una célula pluripotente.

En un aspecto, la divulgación proporciona un método para el análisis de múltiples tipos de células que comprende: escindir al menos un ácido nucleico diana de dos o más células usando un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, para crear dos ácidos nucleicos diana escindidos, insertar un polinucleótido donante diferente en cada uno de los ácidos nucleicos diana escindidos, y analizar las dos o más células. En algunos aspectos, el análisis comprende analizar simultáneamente las dos o más células. En algunos aspectos, el análisis comprende determinar una secuencia del ácido nucleico diana. En algunos aspectos, el análisis comprende comparar las dos o más células. En algunos aspectos, el análisis comprende determinar un genotipo de las dos o más células. En algunos aspectos, la célula es una célula diferenciada. En algunos aspectos, la célula es una célula desdiferenciada. En algunos aspectos, la célula es una célula madre. En algunos aspectos, la célula es una célula madre pluripotente. En algunos aspectos, la célula es una estirpe celular eucariota. En algunos aspectos, la célula es una estirpe celular primaria. En algunos aspectos, la célula es una estirpe celular derivada del paciente. En algunos aspectos, se inserta una pluralidad de polinucleótidos donantes en una pluralidad de ácidos nucleicos diana escindidos de la célula.

En un aspecto, la divulgación proporciona una composición que comprende: un ácido nucleico dirigido a ácido nucleico diseñado que comprende una extensión de hibridación de 3' y un polinucleótido donante, en donde el polinucleótido donante se hibrida con la extensión de hibridación de 3'. En algunos aspectos, la extensión de hibridación de 3' está adaptada para hibridarse con al menos 5 nucleótidos del 3' del polinucleótido donante. En algunos aspectos, la extensión de hibridación de 3' está adaptada para hibridarse con al menos 5 nucleótidos del 5' del polinucleótido donante. En algunos aspectos, la extensión de hibridación de 3' está adaptada para hibridarse con al menos 5 nucleótidos adyacentes en el polinucleótido donante. En algunos aspectos, la extensión de hibridación de 3' está adaptada para hibridarse con todo el polinucleótido donante. En algunos aspectos, la extensión de hibridación de 3' comprende una plantilla de transcripción inversa. En algunos aspectos, la plantilla de transcripción inversa está adaptada para ser transcrita inversamente por una transcriptasa inversa. En algunos aspectos, la composición comprende además un polinucleótido de ADN transcrito inversamente. En algunos aspectos, el polinucleótido de ADN transcrito inversamente está adaptado a hibridarse con la plantilla de transcripción inversa. En algunos aspectos, el polinucleótido donante es ADN. En algunos aspectos, la extensión de hibridación de 3' es ARN. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado aislado. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico diseñado es un ácido nucleico dirigido a ácido nucleico diseñado recombinante.

En un aspecto, la divulgación proporciona un método de introducción de un polinucleótido donante en un ácido nucleico diana que comprende: poner en contacto el ácido nucleico diana con una composición que comprende: un ácido

nucleico dirigido a ácido nucleico diseñado que comprende una extensión de hibridación de 3' y un polinucleótido donante, en donde el polinucleótido donante se hibrida con la extensión de hibridación de 3'. En algunos aspectos, el método comprende además escindir el ácido nucleico diana para producir un ácido nucleico diana escindido. En algunos aspectos, la escisión se realiza mediante un polipéptido dirigido a un sitio. En algunos aspectos, el método comprende además insertar el polinucleótido donante en el ácido nucleico diana escindido.

En un aspecto, la divulgación proporciona una composición que comprende: una proteína efectora y un ácido nucleico, en donde el ácido nucleico comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y una secuencia no nativa, en donde el ácido nucleico está adaptado a unirse a la proteína efectora. En algunos aspectos, la composición comprende además un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa Cas9, en donde el ácido nucleico se une al polipéptido. En algunos aspectos, el polipéptido comprende al menos un 60 % de identidad de secuencia de aminoácidos en un dominio nucleasa con un dominio nucleasa Cas9. En algunos aspectos, el polipéptido es Cas9. En algunos aspectos, el ácido nucleico comprende además una secuencia enlazadora, en donde la secuencia enlazadora une la secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y la secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia no nativa se encuentra en una posición del ácido nucleico seleccionada del grupo que consiste en: un extremo 5' y un extremo 3', o cualquier combinación de los mismos. En algunos aspectos, el ácido nucleico comprende dos moléculas de ácido nucleico. En algunos aspectos, el ácido nucleico comprende una sola molécula continua de ácido nucleico. En algunos aspectos, la secuencia no nativa comprende una secuencia de unión a proteínas de unión a ARN de CRISPR. En algunos aspectos, la secuencia no nativa comprende una secuencia de unión seleccionada del grupo que consiste en: una secuencia de unión a ARN de Cas5, una secuencia de unión a ARN de Cas6 y una secuencia de unión a ARN de Csy4, o cualquier combinación de las mismas. En algunos aspectos, la proteína efectora comprende una proteína de unión a ARN de CRISPR. En algunos aspectos, la proteína efectora comprende al menos un 15 % de identidad de secuencia de aminoácidos con una proteína seleccionada del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, un dominio de unión a ARN de la proteína efectora comprende al menos un 15 % de identidad de secuencia de aminoácidos con un dominio de unión a ARN de una proteína seleccionada del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, la proteína efectora se selecciona del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, la proteína efectora comprende además una o más secuencias no nativas. En algunos aspectos, la secuencia no nativa confiere una actividad enzimática a la proteína efectora. En algunos aspectos, la actividad enzimática se selecciona del grupo que consiste en: actividad de metiltransferasa, actividad de desmetilasa, actividad de acetilación, actividad de desacetilación, actividad de ubiquitinación, actividad de desubiquitinación, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad de formación de dímeros de pirimidina, actividad de transposasa, actividad de recombinasa, actividad de polimerasa, actividad de ligasa, actividad de helicasa, actividad de fotoliasa o actividad de glucosilasa, actividad de acetiltransferasa, actividad de desacetilasa, actividad de quinasa, actividad de fosfatasa, actividad de ubiquitina ligasa, actividad desubiquitinante, actividad de adenilación, actividad de desadenilación, actividad de SUMOilación, actividad deSUMOilación, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación, actividad de remodelación, actividad de proteasa, actividad de oxidorreductasa, actividad de transferasa, actividad de hidrolasa, actividad de liasa, actividad de isomerasa, actividad de sintasa, actividad de sintetasa y actividad de desmiristoilación, o cualquier combinación de las mismas. En algunos aspectos, el ácido nucleico es ARN. En algunos aspectos, la proteína efectora comprende una proteína de fusión que comprende una proteína de unión a ARN y una proteína de unión a ADN. En algunos aspectos, la composición comprende además un polinucleótido donante. En algunos aspectos, el polinucleótido donante se une directamente a la proteína de unión al ADN, y en donde la proteína de unión al ARN se une a ácido nucleico que se dirige a ácido nucleico. En algunos aspectos, el extremo 5' del polinucleótido donante se une a la proteína de unión al ADN. En algunos aspectos, el extremo 3' del polinucleótido donante se une a la proteína de unión al ADN. En algunos aspectos, al menos 5 nucleótidos del polinucleótido donante se unen a la proteína de unión al ADN. En algunos aspectos, el ácido nucleico es un ácido nucleico aislado. En algunos aspectos, el ácido nucleico es un ácido nucleico recombinante.

En un aspecto, la divulgación proporciona un método de introducción de un polinucleótido donante en un ácido nucleico diana que comprende: poner en contacto un ácido nucleico diana con un complejo que comprende un polipéptido dirigido a un sitio y una composición que comprende: una proteína efectora y un ácido nucleico, en donde el ácido nucleico comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y una secuencia no nativa, en donde el ácido nucleico está adaptado a unirse a la proteína efectora. En algunos aspectos, el método comprende además escindir el ácido nucleico diana. En algunos aspectos, la escisión es realizada por el polipéptido dirigido a un sitio. En algunos aspectos, el método comprende además insertar el polinucleótido donante en el ácido nucleico diana.

En un aspecto, la divulgación proporciona un método de modulación de un ácido nucleico diana que comprende: poner en contacto un ácido nucleico diana con uno o más complejos, comprendiendo cada complejo un polipéptido dirigido a un sitio y una composición que comprende: una proteína efectora y un ácido nucleico, en donde el ácido nucleico comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y una secuencia no nativa, en donde

el ácido nucleico está adaptado a unirse a la proteína efectora y modular el ácido nucleico diana. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9. En algunos aspectos, la modulación es realizada por la proteína efectora. En algunos aspectos, la modulación comprende una actividad seleccionada del grupo que consiste en: actividad de metiltransferasa, actividad de desmetilasa, actividad de acetilación, actividad de desacetilación, actividad de ubiquitinación, actividad de desubiquitinación, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad de formación de dímeros de pirimidina, actividad de transposasa, actividad de recombinasa, actividad de polimerasa, actividad de ligasa, actividad de helicasa, actividad de fotoliasa o actividad de glucosilasa, actividad de acetiltransferasa, actividad de desacetilasa, actividad de quinasa, actividad de fosfatasa, actividad de ubiquitina ligasa, actividad desubiquitinante, actividad de adenilación, actividad de desadenilación, actividad de SUMOilación, actividad deSUMOilación, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación, actividad de remodelación, actividad de proteasa, actividad de oxidorreductasa, actividad de transferasa, actividad de hidrolasa, actividad de liasa, actividad de isomerasa, actividad de sintasa, actividad de sintetasa y actividad de desmiristoilación, o cualquier combinación de las mismas. En algunos aspectos, la proteína efectora comprende una o más proteínas efectoras.

En un aspecto, la divulgación proporciona un método para detectar si dos complejos están próximos entre sí, que comprende: poner en contacto un primer ácido nucleico diana con un primer complejo, en donde el primer complejo comprende un primer polipéptido dirigido a un sitio, un primer ácido nucleico dirigido a ácido nucleico modificado y una primera proteína efectora, en donde la proteína efectora está adaptada para unirse a ácido nucleico dirigido a ácido nucleico modificado, y en donde la primera proteína efectora comprende una secuencia no nativa que comprende una primera parte de un sistema dividido, y poner en contacto un segundo ácido nucleico diana con un segundo complejo, en donde el segundo complejo comprende un segundo polipéptido dirigido a un sitio, un segundo ácido nucleico dirigido a ácido nucleico modificado y una segunda proteína efectora, en donde la proteína efectora está adaptada para unirse a ácido nucleico dirigido a ácido nucleico modificado, y en donde la segunda proteína efectora comprende una secuencia no nativa que comprende una segunda parte de un sistema dividido. En algunos aspectos, el primer ácido nucleico diana y el segundo ácido nucleico diana están en el mismo polímero polinucleotídico. En algunos aspectos, el sistema dividido comprende dos o más fragmentos de proteínas que individualmente no son activos, pero, cuando se forman en un complejo, dan lugar a un complejo proteico activo. En algunos aspectos, el método comprende además detectar una interacción entre la primera parte y la segunda parte. En algunos aspectos, la detección indica que el primer y el segundo complejo están cerca uno del otro. En algunos aspectos, el polipéptido dirigido a un sitio está adaptado a no poder escindir el ácido nucleico diana. En algunos aspectos, la detección comprende determinar la existencia de un hecho de movilidad genética. En algunos aspectos, el hecho de movilidad genética comprende una translocación. En algunos aspectos, antes del hecho de movilidad genética, las dos partes del sistema dividido no interactúan. En algunos aspectos, después del hecho de movilidad genética, las dos partes del sistema dividido interactúan. En algunos aspectos, el hecho de movilidad genética es una translocación entre un BCR y un gen Abl. En algunos aspectos, la interacción activa el sistema dividido. En algunos aspectos, la interacción indica que los ácidos nucleicos diana unidos por los complejos están muy juntos entre sí. En algunos aspectos, el sistema dividido se selecciona del grupo que consiste en: sistema de GFP dividido, un sistema de ubiquitina dividido, un sistema de factor de transcripción dividido y un sistema de marcador de afinidad dividido, o cualquier combinación de los mismos. En algunos aspectos, el sistema dividido comprende un sistema de GFP dividido. En algunos aspectos, la detección indica un genotipo. En algunos aspectos, el método comprende además: determinar un curso de tratamiento para una enfermedad basada en el genotipo. En algunos aspectos, el método comprende además tratar la enfermedad. En algunos aspectos, el tratamiento comprende administrar un fármaco. En algunos aspectos, el tratamiento comprende administrar un complejo que comprende un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio, en donde el complejo puede modificar un elemento genético que participe en la enfermedad. En algunos aspectos, la modificación se selecciona del grupo que consiste en: añadir una secuencia de ácido nucleico al elemento genético, sustituir una secuencia de ácido nucleico en el elemento genético, y eliminar una secuencia de ácido nucleico del elemento genético, o cualquier combinación de los mismos. En algunos aspectos, el método comprende además: comunicar el genotipo de un cuidador a un paciente. En algunos aspectos, la comunicación comprende comunicarse desde un sistema de memoria de almacenamiento a un ordenador remoto. En algunos aspectos, la detección diagnostica una enfermedad. En algunos aspectos, el método comprende además: comunicar el diagnóstico de un cuidador a un paciente. En algunos aspectos, la detección indica la presencia de polimorfismo de un solo nucleótido (SNP). En algunos aspectos, el método comprende además: comunicar la existencia de un hecho de movilidad genética de un cuidador a un paciente. En algunos aspectos, la comunicación comprende comunicarse desde un sistema de memoria de almacenamiento a un ordenador remoto. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 20 % de identidad de secuencia de aminoácidos con Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 60 % de identidad de secuencia de aminoácidos con Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 60 % de identidad de secuencia de aminoácidos en un dominio nucleasa con un dominio nucleasa Cas9. En algunos aspectos, el polipéptido dirigido a un sitio es Cas9. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa. En algunos aspectos, la secuencia no nativa se encuentra en una posición del ácido nucleico dirigido a ácido nucleico modificado seleccionada del grupo que consiste en: un extremo 5' y un extremo 3', o cualquier combinación de los mismos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico modificado comprende dos moléculas de ácido nucleico. En algunos aspectos, el ácido nucleico comprende una sola molécula continua de ácido nucleico que comprende una primera parte que comprende al menos un 50 % de identidad con una repetición

CRISPR sobre 6 nucleótidos contiguos y una segunda parte que comprende al menos un 50 % de identidad con una secuencia de ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, la primera parte y la segunda parte están unidas por un enlazador. En algunos aspectos, la secuencia no nativa comprende una secuencia de unión a proteínas de unión a ARN de CRISPR. En algunos aspectos, la secuencia no nativa comprende una secuencia de unión seleccionada del grupo que consiste en: una secuencia de unión a ARN de Cas5, una secuencia de unión a ARN de Cas6 y una secuencia de unión a ARN de Csy4, o cualquier combinación de las mismas. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico modificado está adaptado a unirse a una proteína efectora. En algunos aspectos, la proteína efectora es una proteína de unión a ARN de CRISPR. En algunos aspectos, la proteína efectora comprende al menos un 15 % de identidad de secuencia de aminoácidos con una proteína seleccionada del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, un dominio de unión a ARN de la proteína efectora comprende al menos un 15 % de identidad de secuencia de aminoácidos con un dominio de unión a ARN de una proteína seleccionada del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, la proteína efectora se selecciona del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es ARN. En algunos aspectos, el ácido nucleico diana es ADN. En algunos aspectos, la interacción comprende formar un marcador de afinidad. En algunos aspectos, la detección comprende capturar el marcador de afinidad. En algunos aspectos, el método comprende además secuenciar ácido nucleico unido al primer y segundo complejo. En algunos aspectos, el método comprende además fragmentar el ácido nucleico antes de la captura. En algunos aspectos, la interacción forma un sistema activado. En algunos aspectos, el método comprende además alterar la transcripción de un primer ácido nucleico diana o de un segundo ácido nucleico diana, en donde la alteración es realizada por el sistema activado. En algunos aspectos, el segundo ácido nucleico diana no se une al primer ácido nucleico diana. En algunos aspectos, La alteración de la transcripción del segundo ácido nucleico diana se realiza en *trans*. En algunos aspectos, la alteración de la transcripción del primer ácido nucleico diana se realiza en *cis*. En algunos aspectos, el primer o segundo ácido nucleico diana se selecciona del grupo que consiste en: un ácido nucleico endógeno y un ácido nucleico exógeno, o cualquier combinación de los mismos. En algunos aspectos, la alteración comprende una transcripción creciente del primer o segundo ácido nucleico diana. En algunos aspectos, el primer o segundo ácido nucleico diana comprende un polinucleótido que codifica uno o más genes que causan la muerte celular. En algunos aspectos, el primer o segundo ácido nucleico diana comprende un polinucleótido que codifica un péptido inductor de lisis celular. En algunos aspectos, el primer o segundo ácido nucleico diana comprende un polinucleótido que codifica un antígeno de reclutamiento de células inmunitarias. En algunos aspectos, el primer o segundo ácido nucleico diana comprende un polinucleótido que codifica uno o más genes implicados en la apóptosis. En algunos aspectos, el uno o más genes implicados en la apóptosis comprende caspasas. En algunos aspectos, el uno o más genes implicados en la apóptosis comprende citocinas. En algunos aspectos, uno o más genes implicados en la apóptosis se seleccionan del grupo que consiste en: factor de necrosis tumoral (TNF), receptor del TNF 1 (TNFR1), receptor del TNF 2 (TNFR2), receptor de Fas, FasL, caspasa-8, caspasa-10, caspasa-3, caspasa-9, caspasa-3, caspasa-6, caspasa-7, Bcl-2 y factor inductor de la apóptosis (AIF), o cualquier combinación de los mismos. En algunos aspectos, el primer o segundo ácido nucleico diana comprende un polinucleótido que codifica uno o más ácidos nucleicos dirigidos a ácido nucleico. En algunos aspectos, el uno o más ácidos nucleicos dirigidos a ácido nucleico se dirigen a una pluralidad de ácidos nucleicos diana. En algunos aspectos, la detección comprende generar datos genéticos. En algunos aspectos, el método comprende además comunicar los datos genéticos desde un sistema de memoria de almacenamiento a un ordenador remoto. En algunos aspectos, los datos genéticos indican un genotipo. En algunos aspectos, los datos genéticos indican la existencia de un hecho de movilidad genética. En algunos aspectos, los datos genéticos indican una ubicación espacial de los genes.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, una proteína efectora que está adaptada para unirse a la secuencia no nativa y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa. En algunos aspectos, la secuencia polinucleotídica está unida operativamente a un promotor. En algunos aspectos, el promotor es un promotor inducible.

En un aspecto, la divulgación proporciona un kit que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia configurada para unirse a una proteína efectora, y un polipéptido dirigido a un sitio. En algunos aspectos, la secuencia polinucleotídica está unida operativamente a un promotor. En algunos aspectos, el promotor es un promotor inducible.

En un aspecto, la divulgación proporciona un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, un polipéptido dirigido a un sitio y una proteína efectora. En algunos aspectos, la secuencia polinucleotídica está unida operativamente a un promotor. En algunos aspectos, el promotor es un promotor inducible.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende una composición que comprende: una proteína efectora y un ácido nucleico, en donde el ácido nucleico comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y una secuencia no nativa, en donde el ácido nucleico está adaptado a unirse a la proteína efectora.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia configurada para unirse a una proteína efectora, y un polipéptido dirigido a un sitio.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, un polipéptido dirigido a un sitio y una proteína efectora.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia configurada para unirse a una proteína efectora, y un polipéptido dirigido a un sitio y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, un polipéptido dirigido a un sitio y una proteína efectora, y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona una composición que comprende: un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana. En algunos aspectos, el módulo de ácido nucleico comprende una primera secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una segunda secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, la composición comprende además una secuencia enlazadora que une la primera y la segunda secuencia. En algunos aspectos, el uno o más módulos de ácido nucleico se hibridan con uno o más ácidos nucleicos diana. En algunos aspectos, el uno o más módulos de ácido nucleico difieren en al menos un nucleótido en una región espaciadora del uno o más módulos de ácido nucleico. En algunos aspectos, el uno o más módulos de ácido nucleico es ARN. En algunos aspectos, El agente de dirección genético multiplexado es ARN. En algunos aspectos, la secuencia no nativa comprende una ribozima. En algunos aspectos, la secuencia no nativa comprende una secuencia de unión a endorribonucleasa. En algunos aspectos, la secuencia de unión a endorribonucleasa se encuentra en un extremo 5' del módulo de ácido nucleico. En algunos aspectos, la secuencia de unión a endorribonucleasa se encuentra en un extremo 3' del módulo de ácido nucleico. En algunos aspectos, la secuencia de unión a endorribonucleasa está adaptada ser unida por una endorribonucleasa CRISPR. En algunos aspectos, la secuencia de unión a endorribonucleasa está adaptada para ser unida por una endorribonucleasa que comprende un dominio RAMP. En algunos aspectos, la secuencia de unión a endorribonucleasa está adaptada para ser unida por una endorribonucleasa seleccionada del grupo que consiste en: una endorribonucleasa miembro de la superfamilia de Cas5 y una endorribonucleasa miembro de la superfamilia de Cas6, o cualquier combinación de las mismas. En algunos aspectos, la secuencia de unión a endorribonucleasa está adaptada para ser unida por una endorribonucleasa que comprende al menos un 15 % de identidad de secuencia de aminoácidos con una proteína seleccionada del grupo que consiste en: Csy4, Cas5 y Cas6. En algunos aspectos, la secuencia de unión a endorribonucleasa está adaptada para ser unida por una endorribonucleasa que comprende al menos un 15 % de identidad de secuencia de aminoácidos con un dominio nucleasa de una proteína seleccionada del grupo que consiste en: Csy4, Cas5 y Cas6. En algunos aspectos, la secuencia de unión a endorribonucleasa comprende una horquilla. En algunos aspectos, la horquilla comprende al menos 4 nucleótidos consecutivos en una estructura de tallo-lazo. En algunos aspectos, la secuencia de unión a endorribonucleasa comprende al menos un 60 % de identidad con una secuencia seleccionada del grupo que consiste en:

5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3';
5'-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3';
5'-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3';
5'-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3';
5'-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3';
5'-GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3';
5'-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3';
5'-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3'.
5'-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3';
5'-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3';
5'-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3';
5'-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3';
5'-GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3';
5'-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3';
5'-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3',
5'-GUCGCCCCCCACGCGGGGGCGUGGAUUGAAAC-3';
5'-CC AGCCGCCUUCGGGCGGCUGUGUGUUGAAAC-3';
5'-GUCGCACUCUACAUGAGUGCGUGGAUUGAAAU-3';
5'-UGUCGCACCUUAUAUAGGUGCGUGGAUUGAAAU-3 y
5'-GUCGCGCCCCGCAUGGGGCGCGUGGAUUGAAA-3' o cualquier combinación de las mismas. En algunos aspectos, el uno o más módulos de ácido nucleico están adaptados para ser unidos por diferentes endorribonucleasas. En algunos aspectos, el agente de dirección genético multiplexado es un agente de dirección genético multiplexado aislado. En algunos aspectos, el agente de dirección genético multiplexado es un agente de dirección genético multiplexado recombinante.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana. En algunos aspectos, la secuencia polinucleotídica está unida operativamente a un promotor. En algunos aspectos, el promotor es un promotor inducible.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un vector que comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana.

En un aspecto, la divulgación proporciona un kit que comprende un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un método de generación de un ácido nucleico, en donde el ácido nucleico se une a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y se hibrida con un ácido nucleico diana que comprende: introducir un agente de dirección genético

multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana en una célula hospedadora, procesar el agente de dirección genético multiplexado en uno o más módulos de ácido nucleico, y poner en contacto los uno o más módulos de ácido nucleico procesados con uno o más ácidos nucleicos diana en la célula. En algunos aspectos, el método comprende además escindir el ácido nucleico diana. En algunos aspectos, el método comprende además modificar el ácido nucleico diana. En algunos aspectos, la modificación comprende alterar la transcripción del ácido nucleico diana. En algunos aspectos, la modificación comprende además insertar un polinucleótido donante en el ácido nucleico diana.

En un aspecto, la divulgación proporciona un polipéptido dirigido a un sitio modificado que comprende: un primer dominio nucleasa, un segundo dominio nucleasa y un dominio nucleasa insertado. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad con un dominio nucleasa de Cas9. En algunos aspectos, el primer dominio nucleasa comprende un dominio nucleasa seleccionado del grupo que consiste en: un dominio HNH y un dominio RuvC, o cualquier combinación de los mismos. En algunos aspectos, el segundo dominio nucleasa comprende un dominio nucleasa seleccionado del grupo que consiste en: un dominio HNH y un dominio RuvC, o cualquier combinación de los mismos. En algunos aspectos, el dominio nucleasa insertado comprende un dominio HNH. En algunos aspectos, el dominio nucleasa insertado comprende un dominio RuvC. En algunos aspectos, el dominio nucleasa insertado es N-terminal al primer dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado es N-terminal al segundo dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado es C-terminal al primer dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado es C-terminal al segundo dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado está en tándem con el primer dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado está en tándem con el segundo dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado está adaptado a escindir un ácido nucleico diana en un sitio diferente al primer o segundo dominio nucleasa. En algunos aspectos, el dominio nucleasa insertado está adaptado a escindir un ARN en un híbrido de ADN-ARN. En algunos aspectos, el dominio nucleasa insertado está adaptado a escindir un ADN en un híbrido de ADN-ARN. En algunos aspectos, el dominio nucleasa insertado está adaptado a aumentar la especificidad de unión del polipéptido dirigido a un sitio modificado con un ácido nucleico diana. En algunos aspectos, el dominio nucleasa insertado se adapta para aumentar la fuerza de unión del polipéptido dirigido a un sitio modificado con un ácido nucleico diana.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado que comprende: un primer dominio nucleasa, un segundo dominio nucleasa y un dominio nucleasa insertado.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio modificado que comprende: un primer dominio nucleasa, un segundo dominio nucleasa y un dominio nucleasa insertado, y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo motivo adyacente al protoespaciador en comparación con un polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio se modifica mediante una modificación seleccionada del grupo que consiste en: una adición de aminoácido, una sustitución de aminoácido, un reemplazo de aminoácido y una eliminación de aminoácido, o cualquier combinación de los mismos. En algunos aspectos, el polipéptido dirigido a un sitio modificado comprende una secuencia no nativa. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a dirigirse al segundo motivo adyacente al protoespaciador con mayor especificidad que el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a dirigirse al segundo motivo adyacente al protoespaciador con una constante de disociación más baja en comparación con el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio modificado se adapta para dirigirse al segundo motivo adyacente al protoespaciador con una constante de disociación más alta en comparación con el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el segundo motivo adyacente al protoespaciador comprende un motivo adyacente al protoespaciador seleccionado del grupo que consiste en: 5'-NGGNG-3', 5'-NNAAAAW-3', 5'-NNNNGATT-3', 5'-GNNNCNNA-3' y 5'-NNNACA-3', o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo motivo adyacente al protoespaciador en comparación con un polipéptido dirigido a un sitio de tipo silvestre.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo motivo adyacente al protoespaciador en comparación con un polipéptido dirigido a un sitio de tipo silvestre y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio

modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio se modifica mediante una modificación seleccionada del grupo que consiste en: una adición de aminoácido, una sustitución de aminoácido, un reemplazo de aminoácido y una eliminación de aminoácido, o cualquier combinación de los mismos. En algunos aspectos, el polipéptido dirigido a un sitio modificado comprende una secuencia no nativa. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a dirigirse al segundo ácido nucleico dirigido a ácido nucleico con una especificidad superior en comparación con el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a dirigirse al segundo ácido nucleico dirigido a ácido nucleico con una constante de disociación más baja en comparación con el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a dirigirse al segundo ácido nucleico dirigido a ácido nucleico con una constante de disociación más alta en comparación con el polipéptido dirigido a un sitio de tipo silvestre. En algunos aspectos, el polipéptido dirigido a un sitio se dirige a una parte de ARNcrtra del segundo ácido nucleico dirigido a ácido nucleico.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio de tipo silvestre.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio modificado, en donde el polipéptido se modifica de modo que se adapta para dirigirse a un segundo ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio de tipo silvestre y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente en comparación con la SEQ ID: 8. En algunos aspectos, la composición está configurada para escindir un ácido nucleico diana.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un parche muy básico en comparación con la SEQ ID: 8. En algunos aspectos, la composición está configurada para escindir un ácido nucleico diana.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un dominio de tipo polimerasa en comparación con la SEQ ID: 8. En algunos aspectos, la composición está configurada para escindir un ácido nucleico diana.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de las mismas.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de las mismas.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de las mismas y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso. En algunos aspectos, el kit comprende además un ácido nucleico dirigido a ácido nucleico.

En un aspecto, la divulgación proporciona a una célula modificada genéticamente un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de las mismas.

En un aspecto, la divulgación proporciona un método para el diseño del genoma que comprende: poner en contacto un ácido nucleico diana con un complejo, en donde el complejo comprende un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de las mismas, y un ácido nucleico dirigido a ácido nucleico, y modificar el ácido nucleico diana. En algunos aspectos, la puesta en contacto comprende poner en contacto el complejo con un motivo adyacente al protoespaciador en el ácido nucleico diana. En algunos aspectos, la puesta en contacto comprende poner en contacto el complejo con una secuencia de ácido nucleico diana más larga en comparación con un polipéptido dirigido a un sitio no modificado. En algunos aspectos, la modificación comprende escindir el ácido nucleico diana. En algunos aspectos, el ácido nucleico diana comprende ARN. En algunos aspectos, el ácido nucleico diana comprende ADN. En algunos aspectos, la modificación comprende escindir la cadena de ARN de un ARN y ADN hibridado. En algunos aspectos, la modificación comprende escindir la cadena de ADN de un ARN

y ADN hibridado. En algunos aspectos, la modificación comprende insertar en el ácido nucleico diana un polinucleótido donante, una parte de un polinucleótido donante, una copia de un polinucleótido donante, o una parte de una copia de un polinucleótido donante, o cualquier combinación de las mismas. En algunos aspectos, la modificación comprende modificar la actividad de transcripción del ácido nucleico diana. En algunos aspectos, la modificación comprende una eliminación de uno o más nucleótidos del ácido nucleico diana.

En un aspecto, la divulgación proporciona una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende un dominio nucleasa modificado en comparación con la SEQ ID: 8. En algunos aspectos, la composición está configurada para escindir un ácido nucleico diana. En algunos aspectos, el dominio nucleasa modificado comprende un dominio nucleasa de dominio RuvC. En algunos aspectos, el dominio nucleasa modificado comprende un dominio nucleasa de HNH. En algunos aspectos, el dominio nucleasa modificado comprende la duplicación de un dominio nucleasa de HNH. En algunos aspectos, el dominio nucleasa modificado se adapta para aumentar la especificidad de la secuencia de aminoácidos por un ácido nucleico diana en comparación con un polipéptido dirigido a un sitio no modificado. En algunos aspectos, el dominio nucleasa modificado se adapta para aumentar la especificidad de la secuencia de aminoácidos por un ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio no modificado. En algunos aspectos, el dominio nucleasa modificado comprende una modificación seleccionada del grupo que consiste en: una adición de aminoácido, una sustitución de aminoácido, un reemplazo de aminoácido y una eliminación de aminoácido, o cualquier combinación de los mismos. En algunos aspectos, el dominio nucleasa modificado comprende una secuencia no nativa insertada. En algunos aspectos, la secuencia no nativa confiere una actividad enzimática al polipéptido dirigido a un sitio modificado. En algunos aspectos, la actividad enzimática se selecciona del grupo que consiste en: actividad de nucleasa, actividad de metilasa, actividad de acetilasa, actividad de desmetilasa, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad de formación de dímeros de pirimidina, actividad de integrasa, actividad de transposasa, actividad de recombinasa, actividad de polimerasa, actividad de ligasa, actividad de helicasa, actividad de fotoliasa o actividad de glucosilasa, actividad de acetiltransferasa, actividad de desacetilasa, actividad de quinasa, actividad de fosfatasa, actividad de ubiquitina ligasa, actividad desubiquitinante, actividad de adenilación, actividad de desadenilación, actividad de SUMOilación, actividad deSUMOilación, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación, actividad de remodelación, actividad de proteasa, actividad de oxidorreductasa, actividad de transferasa, actividad de hidrolasa, actividad de liasa, actividad de isomerasa, actividad de sintasa, actividad de sintetasa y actividad de desmiristoilación, o cualquier combinación de las mismas. En algunos aspectos, la actividad enzimática está adaptada para modular la transcripción de un ácido nucleico diana. En algunos aspectos, el dominio nucleasa modificado está adaptado a permitir la unión de la secuencia de aminoácidos a una secuencia de motivo adyacente al protoespaciador que es diferente de una secuencia de motivo adyacente al protoespaciador con la que un polipéptido dirigido a un sitio no modificado está adaptado a unirse. En algunos aspectos, el dominio nucleasa modificado está adaptado a permitir la unión de la secuencia de aminoácidos a un ácido nucleico dirigido a ácido nucleico que es diferente de un ácido nucleico dirigido a ácido nucleico con el que un polipéptido dirigido a un sitio no modificado está adaptado a unirse. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a unirse a una secuencia de ácido nucleico diana más larga que un polipéptido dirigido a un sitio no modificado. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a escindir ADN bicatenario. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a escindir la cadena de ARN de un ARN y un ADN hibridado. En algunos aspectos, el polipéptido dirigido a un sitio modificado está adaptado a escindir la cadena de ADN de un ARN y un ADN hibridado. En algunos aspectos, la composición comprende además un ácido nucleico dirigido a ácido nucleico modificado, en donde la modificación del polipéptido dirigido a un sitio está adaptada a permitir que el polipéptido dirigido a un sitio se una a ácido nucleico dirigido a ácido nucleico modificado. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico modificado y el polipéptido dirigido a un sitio modificado comprenden mutaciones compensatorias.

En un aspecto, la divulgación proporciona un método para enriquecer un ácido nucleico diana para la secuenciación, que comprende: poner en contacto un ácido nucleico diana con un complejo que comprende un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio, enriquecer el ácido nucleico diana usando el complejo y determinar una secuencia del ácido nucleico diana. En algunos aspectos, el método no comprende una etapa de amplificación. En algunos aspectos, el método comprende además analizar la secuencia del ácido nucleico diana. En algunos aspectos, el método comprende además fragmentar el ácido nucleico diana antes del enriquecimiento. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende ARN. En algunos aspectos, el método del ácido nucleico dirigido a ácido nucleico comprende dos moléculas de ARN. En algunos aspectos, en el método, una parte de cada una de las dos moléculas de ARN se hibridan entre sí. En algunos aspectos, en el método, una de las dos moléculas de ARN comprende una secuencia de repetición CRISPR. En algunos aspectos, la secuencia de repetición CRISPR es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de repetición CRISPR comprende al menos un 60 % de identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la una de las dos moléculas de ARN comprende una secuencia de ARNcrtra. En algunos aspectos, la secuencia de ARNcrtra es homóloga a un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de ARNcrtra comprende al menos un 60 % de identidad con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es un ácido nucleico guía doble. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende una molécula de ARN continua, en donde la molécula de ARN continua comprende además dos dominios y un enlazador. En algunos aspectos, una parte de cada uno de los dos dominios de la molécula de ARN continua se hibrida entre sí. En algunos aspectos, la molécula de ARN continua comprende

una secuencia de repetición CRISPR. En algunos aspectos, la secuencia de repetición CRISPR es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de repetición CRISPR comprende al menos un 60 % de identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la molécula de ARN continua comprende una secuencia de ARNcrtra. En algunos aspectos, la secuencia de ARNcrtra es homóloga a un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de ARNcrtra comprende al menos un 60 % de identidad con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es un ácido nucleico guía sencillo. En algunos aspectos, la puesta en contacto comprende hibridar una parte del ácido nucleico dirigido a ácido nucleico con una parte del ácido nucleico diana. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico se hibrida con el ácido nucleico diana sobre una región que comprende 6-20 nucleótidos. En algunos aspectos, el polipéptido dirigido a un sitio comprende Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 20 % de homología con un dominio nucleasa de Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 60 % de homología con Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende un dominio nucleasa diseñado, en donde el dominio nucleasa comprende una actividad nucleasa reducida en comparación con un polipéptido dirigido a un sitio que comprende un dominio nucleasa no diseñado. En algunos aspectos, el polipéptido dirigido a un sitio introduce una rotura monocatenaria en el ácido nucleico diana. En algunos aspectos, el dominio nucleasa diseñado comprende la mutación de un ácido aspártico conservado. En algunos aspectos, el dominio nucleasa diseñado comprende una mutación D10A. En algunos aspectos, el dominio nucleasa diseñado comprende la mutación de una histidina conservada. En algunos aspectos, el dominio nucleasa diseñado comprende una mutación H840A. En algunos aspectos, el polipéptido dirigido a un sitio comprende un marcador de afinidad. En algunos aspectos, el marcador de afinidad se encuentra en el extremo N del polipéptido dirigido a un sitio, el extremo C del polipéptido dirigido a un sitio, una región accesible a la superficie o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad se selecciona de un grupo que comprende: biotina, FLAG, His6x, His9x y una proteína fluorescente, o cualquier combinación de los mismos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende un marcador de afinidad de ácido nucleico. En algunos aspectos, el marcador de afinidad de ácido nucleico se encuentra en el extremo 5' del ácido nucleico dirigido a ácido nucleico, el extremo 3' del ácido nucleico dirigido a ácido nucleico, una región accesible a la superficie o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad de ácido nucleico se selecciona del grupo que comprende una molécula pequeña, marcador fluorescente, un marcador radiactivo, o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad de ácido nucleico comprende una secuencia que está configurada para unirse a Csy4, Cas5, Cas6 o cualquier combinación de las mismas. En algunos aspectos, el marcador de afinidad de ácido nucleico comprende un 50 % de identidad con 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3'. En algunos aspectos, el método comprende además diagnosticar una enfermedad y tomar una decisión de tratamiento específica del paciente, o cualquier combinación de las mismas. En algunos aspectos, la determinación comprende determinar un genotipo. En algunos aspectos, el método comprende además comunicar la secuencia desde un sistema de memoria de almacenamiento a un ordenador remoto. En algunos aspectos, el enriquecimiento comprende poner en contacto un marcador de afinidad del complejo con un agente de captura. En algunos aspectos, el agente de captura comprende un anticuerpo. En algunos aspectos, el agente de captura comprende un soporte sólido. En algunos aspectos, el agente de captura se selecciona del grupo que comprende: Csy4, Cas5 y Cas6. En algunos aspectos, el agente de captura comprende una actividad enzimática reducida en ausencia de imidazol. En algunos aspectos, el agente de captura comprende un dominio enzimático activable, en donde el dominio enzimático activable se activa al entrar en contacto con imidazol. En algunos aspectos, el agente de captura es un miembro de la familia Cas6. En algunos aspectos, el agente de captura comprende un marcador de afinidad. En algunos aspectos, el agente de captura comprende una endorribonucleasa condicionalmente enzimáticamente inactiva que comprende una mutación en un dominio nucleasa. En algunos aspectos, la mutación de una histidina conservada. En algunos aspectos, la mutación comprende una mutación H29A. En algunos aspectos, el ácido nucleico diana está unido al complejo. En algunos aspectos, el ácido nucleico diana es un ácido nucleico escindido que no está unido al complejo. En algunos aspectos, se pone en contacto una pluralidad de complejos con una pluralidad de ácidos nucleicos diana. En algunos aspectos, la pluralidad de ácidos nucleicos diana difieren en al menos un nucleótido. En algunos aspectos, la pluralidad de complejos comprende una pluralidad de ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido.

En un aspecto, la divulgación proporciona un método para escindir un ácido nucleico que comprende: poner en contacto un ácido nucleico diana con dos o más complejos, en donde cada complejo comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, y escindir el ácido nucleico diana, en donde la escisión produce un ácido nucleico diana escindido. En algunos aspectos, la escisión se realiza mediante un dominio nucleasa del polipéptido dirigido a un sitio. En algunos aspectos, el método no comprende amplificación. En algunos aspectos, el método comprende además enriquecer el ácido nucleico diana escindido. En algunos aspectos, el método comprende además secuenciar el ácido nucleico diana escindido. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es ARN. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende dos moléculas de ARN. En algunos aspectos, una parte de cada una de las dos moléculas de ARN se hibridan entre sí. En algunos aspectos, una de las dos moléculas de ARN comprende una secuencia de repetición CRISPR. En algunos aspectos, la secuencia de repetición CRISPR comprende una secuencia que es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de repetición CRISPR comprende una secuencia que tiene al menos un 60 % de identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, una de las dos moléculas de ARN comprende una secuencia de ARNcrtra. En algunos aspectos, la secuencia de ARNcrtra es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de ARNcrtra comprende al menos un 60 % de

identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es un ácido nucleico guía doble. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende una molécula de ARN continua, en donde la molécula de ARN continua comprende además dos dominios y un enlazador. En algunos aspectos, una parte de cada uno de los dos dominios de la molécula de ARN continua se hibrida entre sí. En algunos aspectos, la molécula de ARN continua comprende una secuencia de repetición CRISPR. En algunos aspectos, la secuencia de repetición CRISPR es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de repetición CRISPR comprende al menos un 60 % de identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la molécula de ARN continua comprende una secuencia de ARNcrtra. En algunos aspectos, la secuencia de ARNcrtra es homóloga a un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, la secuencia de ARNcrtra comprende al menos un 60 % de identidad con un ARNcr sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico es un ácido nucleico guía sencillo. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico se hibrida con un ácido nucleico diana. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico se hibrida con un ácido nucleico diana sobre una región, en donde la región comprende al menos 6 nucleótidos y como máximo 20 nucleótidos. En algunos aspectos, el polipéptido dirigido a un sitio es Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende un polipéptido que comprende al menos un 20 % de homología con un dominio nucleasa de Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende un polipéptido que comprende al menos un 60 % de homología con Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende un marcador de afinidad. En algunos aspectos, el marcador de afinidad se encuentra en el extremo N del polipéptido dirigido a un sitio, el extremo C del polipéptido dirigido a un sitio, una región accesible a la superficie o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad se selecciona de un grupo que comprende: biotina, FLAG, His6x, His9x y una proteína fluorescente, o cualquier combinación de los mismos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende un marcador de afinidad de ácido nucleico. En algunos aspectos, el marcador de afinidad de ácido nucleico se encuentra en el extremo 5' del ácido nucleico dirigido a ácido nucleico, el extremo 3' del ácido nucleico dirigido a ácido nucleico, una región accesible a la superficie o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad de ácido nucleico se selecciona del grupo que comprende una molécula pequeña, marcador fluorescente, un marcador radiactivo, o cualquier combinación de los mismos. En algunos aspectos, el marcador de afinidad de ácido nucleico es una secuencia que puede unirse a Csy4, Cas5, Cas6 o cualquier combinación de las mismas. En algunos aspectos, el marcador de afinidad de ácido nucleico comprende un 50 % de identidad con GUUCACUGCCGUAUAGGCAGCUAAGAAA. En algunos aspectos, el ácido nucleico diana es un ácido nucleico escindido que no está unido a los dos o más complejos. En algunos aspectos, los dos o más complejos se ponen en contacto con una pluralidad de ácidos nucleicos diana. En algunos aspectos, la pluralidad de ácidos nucleicos diana difieren en al menos un nucleótido. En algunos aspectos, los dos o más complejos comprenden ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido.

En un aspecto, la divulgación proporciona un método de generación de un banco de ácidos nucleicos diana que comprende: poner en contacto una pluralidad de ácidos nucleicos diana con un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, escindir la pluralidad de ácidos nucleicos diana y purificar la pluralidad de ácidos nucleicos diana para crear el banco de ácidos nucleicos diana. En algunos aspectos, el método comprende además el cribado del banco de ácidos nucleicos diana.

En un aspecto, la divulgación proporciona una composición que comprende: un primer complejo que comprende: un primer polipéptido dirigido a un sitio y un primer ácido nucleico dirigido a ácido nucleico, un segundo complejo que comprende: un segundo polipéptido dirigido a un sitio y un segundo ácido nucleico dirigido a ácido nucleico, en donde, el primer y el segundo ácido nucleico dirigido a ácido nucleico son diferentes. En algunos aspectos, la composición comprende además un ácido nucleico diana, que es unido por el primer o el segundo complejo. En algunos aspectos, el primer polipéptido dirigido a un sitio y el segundo polipéptido dirigido a un sitio son iguales. En algunos aspectos, el primer polipéptido dirigido a un sitio y el segundo polipéptido dirigido a un sitio son diferentes.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica: dos o más ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido y un polipéptido dirigido a un sitio.

En un aspecto, la divulgación proporciona una célula hospedadora modificada genéticamente que comprende: un vector que comprende una secuencia polinucleotídica que codifica: dos o más ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido y un polipéptido dirigido a un sitio.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende una secuencia polinucleotídica que codifica: dos o más ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido, un polipéptido dirigido a un sitio y un tampón adecuado. En algunos aspectos, el kit comprende además: un agente de captura, un soporte sólido, adaptadores de secuenciación y un control positivo, o cualquier combinación de los mismos. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un kit que comprende: un polipéptido dirigido a un sitio que comprende actividad enzimática reducida en comparación con un polipéptido dirigido a un sitio de tipo silvestre, un ácido nucleico dirigido a ácido nucleico y un agente de captura. En algunos aspectos, el kit comprende además: instrucciones de uso.

En algunos aspectos, el kit comprende además un tampón seleccionado del grupo que comprende: un tampón de lavado, un tampón de estabilización, un tampón reconstituyente o un tampón diluyente.

En un aspecto, la divulgación proporciona un método de escisión de un ácido nucleico diana usando dos o más nickasas que comprende: poner en contacto un ácido nucleico diana con un primer complejo y un segundo complejo, en donde el primer complejo comprende una primera nickasa y un primer ácido nucleico dirigido a ácido nucleico, y en donde el segundo complejo comprende una segunda nickasa y un segundo ácido nucleico dirigido a ácido nucleico, en donde el ácido nucleico diana comprende un primer motivo adyacente al protoespaciador en una primera cadena y un segundo motivo adyacente al protoespaciador en una segunda cadena, en donde los primeros ácidos nucleicos dirigidos a ácido nucleico están adaptados a hibridarse con el primer motivo adyacente al protoespaciador, y en donde el segundo ácido nucleico dirigido a ácido nucleico está adaptado a hibridarse con el segundo motivo adyacente al protoespaciador, y mellar la primera y segunda cadena del ácido nucleico diana, en donde la mella genera un ácido nucleico diana escindido. En algunos aspectos, la primera y la segunda nickasa son iguales. En algunos aspectos, la primera y segunda nickasa son diferentes. En algunos aspectos, el primer y el segundo ácido nucleico dirigido a ácido nucleico son diferentes. En algunos aspectos, hay menos de 125 nucleótidos entre el primer motivo adyacente al protoespaciador y el segundo motivo adyacente al protoespaciador. En algunos aspectos, el primer y segundo motivo adyacente al protoespaciador forman parte de la secuencia ngG, donde N es cualquier nucleótido. En algunos aspectos, la primera o segunda nickasa comprende al menos un dominio nucleasa esencialmente inactivo. En algunos aspectos, la primera o segunda nickasa comprende una mutación de un ácido aspártico conservado. En algunos aspectos, la mutación es una mutación D10A. En algunos aspectos, la primera o segunda nickasa comprende una mutación de una histidina conservada. En algunos aspectos, la mutación es una mutación H840A. En algunos aspectos, hay menos de 15 nucleótidos entre el primer y el segundo motivo adyacente al protoespaciador. En algunos aspectos, hay menos de 10 nucleótidos entre el primer y el segundo motivo adyacente al protoespaciador. En algunos aspectos, hay menos de 5 nucleótidos entre el primer y el segundo motivo adyacente al protoespaciador. En algunos aspectos, el primer y el segundo motivo adyacente al protoespaciador son adyacentes entre sí. En algunos aspectos, el mellado comprende el primer mellado de la nickasa de la primera cadena y el segundo mellado de la nickasa de la segunda cadena. En algunos aspectos, el mellado genera un corte final pegajoso. En algunos aspectos, el mellado genera un corte final romo. En algunos aspectos, el método comprende además insertar un polinucleótido donante en el ácido nucleico diana escindido.

En un aspecto, la divulgación proporciona una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico está adaptada a unirse a su sitio de unión a proteínas de unión a ácido nucleico afín. En algunos aspectos, una o más de la pluralidad de proteínas de unión a ácido nucleico comprenden una secuencia no nativa. En algunos aspectos, la secuencia no nativa se encuentra en una posición seleccionada del grupo que consiste en: el extremo N, el extremo C, una región accesible a la superficie o cualquier combinación de los mismos. En algunos aspectos, la secuencia no nativa codifica una señal de localización nuclear. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico están separadas por un enlazador. En algunos aspectos, algunas de la pluralidad de proteínas de unión a ácido nucleico son la misma proteína de unión a ácido nucleico. En algunos aspectos, toda la pluralidad de proteínas de unión a ácido nucleico son la misma proteína de unión a ácido nucleico. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico son diferentes proteínas de unión a ácido nucleico. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico comprende proteínas de unión a ARN. En algunos aspectos, las proteínas de unión a ARN se seleccionan del grupo que consiste en: una endorribonucleasa de sistema de grupos de repeticiones palindrómicas cortas en intervalos regulares de Tipo I, una endorribonucleasa de sistema de grupos de repeticiones palindrómicas cortas en intervalos regulares de Tipo II o una endorribonucleasa de sistema de grupos de repeticiones palindrómicas cortas en intervalos regulares Tipo III, o cualquier combinación de las mismas. En algunos aspectos, las proteínas de unión a ARN se seleccionan del grupo que consiste en: Cas5, Cas6 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico comprende proteínas de unión a ADN. En algunos aspectos, el sitio de unión a la proteína de unión a ácido nucleico está configurado para unirse a una proteína de unión a ácido nucleico seleccionada del grupo que consiste en: proteína de unión a ácido nucleico del sistema de grupos de repeticiones palindrómicas cortas en intervalos regulares de Tipo I, Tipo II y Tipo III, o cualquier combinación de los mismos. En algunos aspectos, el sitio de unión a la proteína de unión a ácido nucleico está configurado para unirse a una proteína de unión a ácido nucleico seleccionada del grupo que consiste en: Cas6, Cas5 y Csy4, o cualquier combinación de las mismas. En algunos aspectos, algunas de la pluralidad de moléculas de ácido nucleico comprenden el mismo sitio de unión de proteína de unión a ácido nucleico. En algunos aspectos, la pluralidad de moléculas de ácido nucleico comprende el mismo sitio de unión de proteína de unión a ácido nucleico. En algunos aspectos, ninguna de la pluralidad de moléculas de ácido nucleico comprenden el mismo sitio de unión de proteína de unión a ácido nucleico. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 20 % de identidad de secuencia con un dominio nucleasa de Cas9. En algunos aspectos, el polipéptido dirigido a un sitio es Cas9. En algunos aspectos, al menos una de las moléculas de ácido nucleico codifica una endorribonucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares. En algunos aspectos, la endorribonucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares comprende al menos un 20 % de similitud de secuencia con

Csy4. En algunos aspectos, la endorribonucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares comprende al menos un 60 % de similitud de secuencia con Csy4. En algunos aspectos, la endorribonucleasa de grupos de repeticiones palindrómicas cortas en intervalos regulares es Csy4. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico comprende actividad enzimática reducida. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse al sitio de unión de proteína de unión a ácido nucleico, pero no puede escindir el sitio de unión de proteína de unión a ácido nucleico. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende dos moléculas de ARN. En algunos aspectos, una parte de cada una de las dos moléculas de ARN se hibridan entre sí. En algunos aspectos, una primera molécula de las dos moléculas de ARN comprende una secuencia que comprende al menos el 60 % de identidad con una secuencia de ARN de grupos de repeticiones palindrómicas cortas en intervalos regulares sobre 8 nucleótidos contiguos, y una segunda molécula de las dos moléculas de ARN comprende una secuencia que comprende al menos el 60 % de identidad con una secuencia de ARN de grupos de repeticiones palindrómicas cortas en intervalos regulares de activación en *trans* sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprende una molécula de ARN continua, en donde la molécula de ARN continua comprende además dos dominios y un enlazador. En algunos aspectos, una parte de los dos dominios de la molécula de ARN continua se hibrida entre sí. En algunos aspectos, una primera parte de la molécula de ARN continua comprende una secuencia que comprende al menos el 60 % de identidad con una secuencia de ARN de grupos de repeticiones palindrómicas cortas en intervalos regulares sobre 8 nucleótidos contiguos, y una segunda parte de la molécula de ARN continua comprende una secuencia que comprende al menos el 60 % de identidad con una secuencia de ARN de grupos de repeticiones palindrómicas cortas en intervalos regulares de activación en *trans* sobre 6 nucleótidos contiguos. En algunos aspectos, el ácido nucleico dirigido a ácido nucleico está adaptado a hibridarse con un ácido nucleico diana sobre 6-20 nucleótidos. En algunos aspectos, la composición está configurada para ser administrada a una célula. En algunos aspectos, la composición está configurada para administrar cantidades iguales de la pluralidad de moléculas de ácido nucleico a una célula. En algunos aspectos, la composición comprende además una molécula de polinucleótido donante, en donde la molécula de polinucleótido donante comprende un sitio de unión de proteínas de unión a ácido nucleico, en donde el sitio de unión está unido por una proteína de unión a ácido nucleico del polipéptido de fusión.

En un aspecto, la divulgación proporciona un método de administración de ácidos nucleicos a una ubicación subcelular en una célula que comprende: introducir en una célula una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica, formar una unidad que comprende un polipéptido dirigido a un sitio traducido de la molécula de ácido nucleico que codifica un polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico, y escindir un ácido nucleico diana, en donde el polipéptido dirigido a un sitio de la unidad escinde el ácido nucleico diana. En algunos aspectos, la pluralidad de proteínas de unión a ácido nucleico se unen a su sitio de unión de proteínas de unión a ácido nucleico afín. En algunos aspectos, una endorribonucleasa escinde uno de los uno o más sitios de unión de proteínas de unión a ácido nucleico. En algunos aspectos, una endorribonucleasa escinde los sitios de unión de proteínas de unión a ácido nucleico del ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico, liberando así el ácido nucleico dirigido a ácido nucleico. En algunos aspectos, la ubicación subcelular se selecciona del grupo que consiste en: la nucleasa, el RE, el aparato de Golgi, las mitocondrias, la pared celular, el lisosoma y el núcleo. En algunos aspectos, la ubicación subcelular es el núcleo.

En un aspecto, la divulgación proporciona un kit que comprende: una secuencia polinucleotídica que codifica una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio; y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica. En algunos aspectos, el vector comprende además un polinucleótido que codifica un promotor. En algunos aspectos, el promotor está unido operativamente al polinucleótido. En algunos aspectos, el promotor es un promotor inducible.

En un aspecto, la divulgación proporciona un organismo modificado genéticamente que comprende un vector que comprende: una secuencia polinucleotídica que codifica una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica.

En un aspecto, la divulgación proporciona un organismo modificado genéticamente que comprende: una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio; y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico está adaptada a unirse a su sitio de unión a proteínas de unión a ácido nucleico afín.

En un aspecto, la divulgación proporciona un kit que comprende: una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico está adaptada a unirse a su sitio de unión a proteínas de unión a ácido nucleico afín y un tampón.

En un aspecto, la divulgación proporciona un kit que comprende: un vector que comprende: una secuencia polinucleotídica que codifica una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio; y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica y un tampón. En algunos aspectos, El kit comprende además instrucciones de uso. En algunos aspectos, el tampón se selecciona del grupo que comprende: un tampón de dilución, un tampón de reconstitución y un tampón de estabilización, o cualquier combinación de los mismos

En un aspecto, la divulgación proporciona un polinucleótido donante que comprende: un elemento genético de interés y un elemento indicador, en donde el elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos. En algunos aspectos, el elemento genético de interés comprende un gen. En algunos aspectos, el elemento genético de interés comprende un ácido nucleico no codificante seleccionado del grupo que consiste en: un microARN, un ARNip y un ARN largo no codificante, o cualquier combinación de los mismos. En algunos aspectos, el elemento genético de interés comprende un gen no codificante. En algunos aspectos, el elemento genético de interés comprende un ácido nucleico no codificante seleccionado del grupo que consiste en: un microARN, un ARNip y un ARN largo no codificante, o cualquier combinación de los mismos. En algunos aspectos, el elemento indicador comprende un gen seleccionado del grupo que consiste en: un gen que codifica una proteína fluorescente, un gen que codifica una proteína quimioluminiscente y un gen de resistencia a antibióticos, o cualquier combinación de los mismos. En algunos aspectos, el elemento indicador comprende un gen que codifica una proteína fluorescente. En algunos aspectos, la proteína fluorescente comprende proteína verde fluorescente. En algunos aspectos, el elemento indicador está unido operativamente a un promotor. En algunos aspectos, el promotor comprende un promotor inducible. En algunos aspectos, el promotor comprende un promotor específico del tejido. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 95 % de identidad de secuencia de aminoácidos sobre 10 aminoácidos para Cas9. En algunos aspectos, el dominio nucleasa se selecciona del grupo que consiste en: un dominio HNH, un dominio de tipo HNH, un dominio RuvC y un dominio de tipo RuvC, o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona un vector de expresión que comprende una secuencia polinucleotídica que codifica un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos.

En un aspecto, la divulgación proporciona una célula modificada genéticamente que comprende un polinucleótido donante que comprende: un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos.

En un aspecto, la divulgación proporciona un kit que comprende: un polinucleótido donante que comprende: un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una secuencia

polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y un tampón. En algunos aspectos, el kit comprende además: un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con Cas9; y un ácido nucleico, en donde el ácido nucleico se une al polipéptido y se hibrida con un ácido nucleico diana. En algunos aspectos, El kit comprende además instrucciones de uso. En algunos aspectos, el kit comprende además un polinucleótido que codifica un polipéptido, en donde el polipéptido comprende al menos un 15 % de identidad de secuencia de aminoácidos con Cas9. En algunos aspectos, el kit comprende además un polinucleótido que codifica un ácido nucleico, en donde el ácido nucleico comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos.

En un aspecto, la divulgación proporciona un método de selección de una célula usando un elemento indicador y escindiendo el elemento indicador de la célula que comprende: poner en contacto el ácido nucleico diana con un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico; escindir el ácido nucleico diana con el polipéptido dirigido a un sitio, para generar un ácido nucleico diana escindido; insertar el polinucleótido donante que comprende un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos en el ácido nucleico diana escindido; y seleccionar la célula basada en el polinucleótido donante para generar una célula seleccionada. En algunos aspectos, la selección comprende seleccionar la célula de un sujeto que está siendo tratado por una enfermedad. En algunos aspectos, la selección comprende seleccionar la célula de un sujeto que está siendo diagnosticado por una enfermedad. En algunos aspectos, después de la selección, la célula comprende el polinucleótido donante. En algunos aspectos, el método comprende además escindir todo, parte o nada del elemento indicador, generando así una segunda célula seleccionada. En algunos aspectos, la escisión comprende poner en contacto el extremo 5' del elemento indicador con un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, en donde el complejo escinde el extremo 5'. En algunos aspectos, la escisión comprende poner en contacto el extremo 3' del elemento indicador con un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, en donde el complejo escinde el extremo 3'. En algunos aspectos, la escisión comprende el contacto del extremo 5' y 3' del elemento indicador con uno o más complejos que comprenden un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, en donde el complejo escinde el extremo 5' y 3'. En algunos aspectos, el método comprende además el cribado de la segunda célula seleccionada. En algunos aspectos, el cribado consiste en observar la ausencia de todo o parte del elemento indicador.

En un aspecto, la divulgación proporciona una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio. En algunos aspectos, la secuencia que está a 5' de un PAM tiene al menos 18 nucleótidos de longitud. En algunos aspectos, la secuencia que está a 5' de un PAM es adyacente al PAM. En algunos aspectos, el PAM comprende 5'-NGG-3'. En algunos aspectos, el primer dúplex está adyacente al espaciador. En algunos aspectos, el dominio P comienza desde 1-5 nucleótidos cadena abajo del dúplex, comprende al menos 4 nucleótidos, y está adaptado a hibridarse con la secuencia seleccionada del grupo que consiste en: una secuencia de motivo adyacente al protoespaciador 5'-NGG-3', una secuencia que comprende al menos un 50 % de identidad con los aminoácidos 1096-1225 de Cas9 de *S. pyogenes*, o cualquier combinación de las mismas. En algunos aspectos, el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad con un dominio nucleasa de Cas9 de *S. pyogenes*. En algunos aspectos, el ácido nucleico es ARN. En algunos aspectos, El ácido nucleico es un ARN en forma de A. En algunos aspectos, el primer dúplex tiene al menos 6 nucleótidos de longitud. En algunos aspectos, los 3 nucleótidos no emparejados de la protuberancia comprenden 5'-AAG-3 '. En algunos aspectos, adyacente a los 3 nucleótidos no emparejados hay un nucleótido que forma un par oscilante con un nucleótido en la segunda cadena del primer dúplex. En algunos aspectos, el polipéptido se une a una región del ácido nucleico seleccionada del grupo que consiste en: el primer dúplex, el segundo dúplex y el dominio P, o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona un método de modificación de un ácido nucleico diana que comprende: poner en contacto un ácido nucleico diana con una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer

dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; y modificar el ácido nucleico diana. En algunos aspectos, el método comprende además entrar en contacto con un polipéptido dirigido a un sitio. En algunos aspectos, la puesta en contacto comprende poner en contacto el espaciador con el ácido nucleico diana. En algunos aspectos, la modificación comprende escindir el ácido nucleico diana para producir un ácido nucleico diana escindido. En algunos aspectos, la escisión es realizada por el polipéptido dirigido a un sitio. En algunos aspectos, el método comprende además insertar un polinucleótido donante en el ácido nucleico diana escindido. En algunos aspectos, la modificación comprende modificar la transcripción del ácido nucleico diana.

En un aspecto, la divulgación proporciona un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio.

En un aspecto, la divulgación proporciona un kit que comprende: una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; y un tampón. En algunos aspectos, el kit comprende además un polipéptido dirigido a un sitio. En algunos aspectos, el kit comprende además un polinucleótido donante. En algunos aspectos, El kit comprende además instrucciones de uso.

En un aspecto, la divulgación proporciona un método de creación de un ácido nucleico dirigido a ácido nucleico diseñado sintéticamente que comprende: diseñar una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio.

En un aspecto, la divulgación proporciona una composición farmacéutica que comprende un ácido nucleico dirigido a ácido nucleico diseñado seleccionado del grupo que consiste en: un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P de dicho ácido nucleico dirigido a ácido nucleico; un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico.

En un aspecto, la divulgación proporciona una composición farmacéutica que comprende una composición seleccionada del grupo que consiste en: una composición que comprende: un ácido nucleico dirigido a ácido nucleico diseñado que comprende una extensión de hibridación de 3' y un polinucleótido donante, en donde dicho polinucleótido donante se hibrida con dicha extensión de hibridación de 3'; una composición que comprende: una proteína efectora y un ácido nucleico, en donde dicho ácido nucleico comprende: al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos y una secuencia no nativa, en donde dicho ácido nucleico está adaptado a unirse a dicha proteína efectora; una composición que comprende: un agente de dirección genético multiplexado, en donde dicho agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde dicho módulo de ácido nucleico comprende una secuencia no nativa, y en donde dicho módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde dicho módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana; una composición que comprende: un polipéptido dirigido a un sitio modificado, en donde dicho polipéptido se modifica de modo que se adapta para dirigirse a un segundo motivo adyacente al protoespaciador en comparación con un polipéptido dirigido a un sitio de tipo silvestre; una composición que comprende: un polipéptido dirigido a un sitio modificado, en donde dicho polipéptido se modifica de modo que se adapta para dirigirse a un segundo ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio de tipo silvestre; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente en comparación con la SEQ ID: 8; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un parche muy básico en comparación con la SEQ ID: 8; una

composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un dominio de tipo polimerasa en comparación con la SEQ ID: 8; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de los mismos; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende un dominio nucleasa modificado en comparación con la SEQ ID: 8; una composición que comprende: un primer complejo que comprende: un primer polipéptido dirigido a un sitio y un primer ácido nucleico dirigido a ácido nucleico, un segundo complejo que comprende: un segundo polipéptido dirigido a un sitio y un segundo ácido nucleico dirigido a ácido nucleico, en donde, dicho primer y dicho segundo ácido nucleico dirigido a ácido nucleico son diferentes; una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de dicha pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de dicha pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde dicho polipéptido de fusión comprende una pluralidad de dichas proteínas de unión a ácido nucleico, en donde dicha pluralidad de proteínas de unión a ácido nucleico está adaptada a unirse a su sitio de unión a proteínas de unión a ácido nucleico afín; y una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde dicho espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde dicho espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM, un primer dúplex, en donde dicho primer dúplex está a 3' de dicho espaciador, una protuberancia, en donde dicha protuberancia comprende al menos 3 nucleótidos no emparejados en dicha primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena de dicho primer dúplex, un enlazador, en donde dicho enlazador une dicha primera cadena y dicha segunda cadena de dicho dúplex y tiene al menos 3 nucleótidos de longitud, un dominio P y un segundo dúplex, en donde dicho segundo dúplex está a 3' de dicho dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona una composición farmacéutica que comprende un polipéptido dirigido a un sitio modificado que comprende: un primer dominio nucleasa, un segundo dominio nucleasa y un dominio nucleasa insertado.

En un aspecto, la divulgación proporciona una composición farmacéutica que comprende un polinucleótido donante que comprende: un elemento genético de interés y un elemento indicador, en donde dicho elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde dichos uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos.

En un aspecto, la divulgación proporciona una composición farmacéutica que comprende un vector seleccionado del grupo que consiste en: un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P de dicho ácido nucleico dirigido a ácido nucleico; un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; y modificar el ácido nucleico diana; un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa; un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia configurada para unirse a una proteína efectora, y un polipéptido dirigido a un sitio; un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, un polipéptido dirigido a un sitio y una proteína efectora; un vector que comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana; un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de los mismos; un vector que comprende una secuencia polinucleotídica que codifica: dos o más ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido; y un polipéptido dirigido a un sitio; un vector que comprende: una secuencia polinucleotídica que codifica una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio; y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica; un vector de expresión que comprende una secuencia polinucleotídica que codifica un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una

secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona un método de tratamiento de una enfermedad que comprende administrar a un sujeto: un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P de dicho ácido nucleico dirigido a ácido nucleico; un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; una composición que comprende: un ácido nucleico dirigido a ácido nucleico diseñado que comprende una extensión de hibridación de 3' y un polinucleótido donante, en donde dicho polinucleótido donante se hibrida con dicha extensión de hibridación de 3'; una composición que comprende: una proteína efectora y un ácido nucleico, en donde dicho ácido nucleico comprende: al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos, al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos y una secuencia no nativa, en donde dicho ácido nucleico está adaptado a unirse a dicha proteína efectora; una composición que comprende: un agente de dirección genético multiplexado, en donde dicho agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde dicho módulo de ácido nucleico comprende una secuencia no nativa, y en donde dicho módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde dicho módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana; una composición que comprende: un polipéptido dirigido a un sitio modificado, en donde dicho polipéptido se modifica de modo que se adapta para dirigirse a un segundo motivo adyacente al protoespaciador en comparación con un polipéptido dirigido a un sitio de tipo silvestre; una composición que comprende: un polipéptido dirigido a un sitio modificado, en donde dicho polipéptido se modifica de modo que se adapta para dirigirse a un segundo ácido nucleico dirigido a ácido nucleico en comparación con un polipéptido dirigido a un sitio de tipo silvestre; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente en comparación con la SEQ ID: 8; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un parche muy básico en comparación con la SEQ ID: 8; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en un dominio de tipo polimerasa en comparación con la SEQ ID: 8; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de los mismos; una composición que comprende: un polipéptido dirigido a un sitio modificado que comprende un dominio nucleasa modificado en comparación con la SEQ ID: 8; una composición que comprende: un primer complejo que comprende: un primer polipéptido dirigido a un sitio y un primer ácido nucleico dirigido a ácido nucleico, un segundo complejo que comprende: un segundo polipéptido dirigido a un sitio y un segundo ácido nucleico dirigido a ácido nucleico, en donde, dicho primer y dicho segundo ácido nucleico dirigido a ácido nucleico son diferentes; una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de dicha pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de dicha pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio y un polipéptido de fusión, en donde dicho polipéptido de fusión comprende una pluralidad de dichas proteínas de unión a ácido nucleico, en donde dicha pluralidad de proteínas de unión a ácido nucleico está adaptada a unirse a su sitio de unión a proteínas de unión a ácido nucleico afín; una composición que comprende: un ácido nucleico que comprende: un espaciador, en donde dicho espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde dicho espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM, un primer dúplex, en donde dicho primer dúplex está a 3' de dicho espaciador, una protuberancia, en donde dicha protuberancia comprende al menos 3 nucleótidos no emparejados en dicha primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena de dicho primer dúplex, un enlazador, en donde dicho enlazador une dicha primera cadena y dicha segunda cadena de dicho dúplex y tiene al menos 3 nucleótidos de longitud, un dominio P y un segundo dúplex, en donde dicho segundo dúplex está a 3' de dicho dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; un polipéptido dirigido a un sitio modificado que comprende: un primer dominio nucleasa, un segundo dominio nucleasa y un dominio nucleasa insertado; un polinucleótido donante que comprende: un elemento genético de interés y un elemento indicador, en donde dicho elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde dichos uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en un dominio P de dicho ácido nucleico dirigido a ácido nucleico;

un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende: una mutación en una región protuberante de un ácido nucleico dirigido a ácido nucleico; y modificar el ácido nucleico diana; un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa; un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia configurada para unirse a una proteína efectora, y un polipéptido dirigido a un sitio; un vector que comprende: una secuencia polinucleotídica que codifica: un ácido nucleico dirigido a ácido nucleico modificado, en donde el ácido nucleico dirigido a ácido nucleico modificado comprende una secuencia no nativa, un polipéptido dirigido a un sitio y una proteína efectora; un vector que comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, en donde el agente de dirección genético multiplexado comprende uno o más módulos de ácido nucleico, en donde el módulo de ácido nucleico comprende una secuencia no nativa, y en donde el módulo de ácido nucleico está configurado para unirse a un polipéptido que comprende al menos un 10 % de identidad de secuencia de aminoácidos con un dominio nucleasa de Cas9 y en donde el módulo de ácido nucleico está configurado para hibridarse a un ácido nucleico diana; un vector que comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio modificado que comprende una modificación en una hélice puente, un parche muy básico, un dominio nucleasa y un dominio polimerasa en comparación con la SEQ ID: 8, o cualquier combinación de los mismos; un vector que comprende una secuencia polinucleotídica que codifica: dos o más ácidos nucleicos dirigidos a ácido nucleico que difieren en al menos un nucleótido; y un polipéptido dirigido a un sitio; un vector que comprende: una secuencia polinucleotídica que codifica una composición que comprende: una pluralidad de moléculas de ácido nucleico, en donde cada molécula de ácido nucleico comprende un sitio de unión de proteína de unión a ácido nucleico, en donde al menos una de la pluralidad de moléculas de ácido nucleico codifica un ácido nucleico dirigido a ácido nucleico, y una de la pluralidad de moléculas de ácido nucleico codifica un polipéptido dirigido a un sitio; y un polipéptido de fusión, en donde el polipéptido de fusión comprende una pluralidad de proteínas de unión a ácido nucleico, en donde la pluralidad de proteínas de unión a ácido nucleico están adaptadas a unirse a sus sitios de unión de proteínas de unión a ácido nucleico afines que administran la composición a la ubicación subcelular de forma estequiométrica; un vector de expresión que comprende una secuencia polinucleotídica que codifica un elemento genético de interés; y un elemento indicador, en donde el elemento indicador comprende una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio, y uno o más ácidos nucleicos, en donde el uno o más ácidos nucleicos comprende una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcr sobre 6 nucleótidos contiguos y una secuencia que comprende al menos un 50 % de identidad de secuencia con un ARNcrtra sobre 6 nucleótidos contiguos; y un vector que comprende una secuencia polinucleotídica que codifica un ácido nucleico que comprende: un espaciador, en donde el espaciador está entre 12-30 nucleótidos, ambos incluidos, y en donde el espaciador está adaptado a hibridarse a una secuencia que está a 5' de un PAM; un primer dúplex, en donde el primer dúplex está a 3' del espaciador; una protuberancia, en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en una primera cadena del primer dúplex y al menos 1 nucleótido no emparejado en una segunda cadena del primer dúplex; un enlazador, en donde el enlazador une la primera cadena y la segunda cadena del dúplex y tiene al menos 3 nucleótidos de longitud; un dominio P; y un segundo dúplex, en donde el segundo dúplex está a 3' del dominio P y está adaptado a unirse a un polipéptido dirigido a un sitio; o cualquier combinación de los mismos. En algunos aspectos, la administración comprende administrar mediante administración vírica. En algunos aspectos, la administración comprende administrar mediante electroporación. En algunos aspectos, la administración comprende administrar mediante administración en nanopartículas. En algunos aspectos, la administración comprende administrar mediante administración en liposomas. En algunos aspectos, la administración comprende administrar mediante un método seleccionado del grupo que consiste en: administración intravenosa, subcutánea, intramuscular, oral, rectal, por aerosol, parenteral, oftálmica, pulmonar, transdérmica, vaginal, ótica, nasal y tópica, o cualquier combinación de las mismas. En algunas realizaciones, los métodos de divulgación se realizan en una célula seleccionada del grupo que consiste en: célula vegetal, célula microbiana y célula fúngica, o cualquier combinación de las mismas.

Breve descripción de los dibujos

Las figuras que no pertenecen a la invención son solo para fines ilustrativos. Las nuevas características de la invención se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención haciendo referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en las que se utilizan los principios de la invención y los dibujos adjuntos de los cuales:

La **Figura 1A** representa un ácido nucleico dirigido a ácido nucleico guía sencillo de la divulgación.
La **Figura 1B** representa una realización a modo de ejemplo de un ácido nucleico dirigido a ácido nucleico guía sencillo de la divulgación.
La **Figura 2** representa un ácido nucleico dirigido a ácido nucleico guía doble de la divulgación.
La **Figura 3** representa un método de enriquecimiento de secuencias de la divulgación que utiliza la escisión del ácido nucleico diana.
La **Figura 4** representa un método de enriquecimiento de secuenciad de la divulgación que utiliza el enriquecimiento del ácido nucleico diana.
La **Figura 5** representa un método de la divulgación para determinar sitios de unión fuera de la diana de un polipéptido dirigido a un sitio utilizando la purificación del polipéptido dirigido a un sitio.

La **Figura 6** representa un método de la divulgación para determinar sitios de unión fuera de la diana de un polipéptido dirigido a un sitio utilizando la purificación del ácido nucleico dirigido a ácido nucleico.
La **Figura 7** ilustra un método de secuenciación basado en una matriz que usa un polipéptido dirigido a un sitio de la divulgación.
La **Figura 8** ilustra un método de secuenciación basado en una matriz que usa un polipéptido dirigido a un sitio de la divulgación, en donde se secuencian productos escindidos.
La **Figura 9** ilustra un método basado en la secuenciación de próxima generación que usa un polipéptido dirigido a un sitio de la divulgación.
La **Figura 10** representa un ejemplo de ácido nucleico dirigido a ácido nucleico guía sencillo marcado.
La **Figura 11** representa un ejemplo de ácido nucleico dirigido a ácido nucleico guía doble marcado.
La **Figura 12** ilustra un método de uso de ácido nucleico dirigido a ácido nucleico marcado con un sistema dividido (por ejemplo, sistema fluorescente dividido).
La **Figura 13** representa algunos datos ilustrativos acerca del efecto de un ácido nucleico dirigido a ácido nucleico marcado en 5' sobre la escisión de ácido nucleico diana.
La **Figura 14** ilustra un ejemplo de ácido nucleico dirigido a ácido nucleico marcado en 5' que comprende una secuencia enlazadora de marcador entre el ácido nucleico dirigido a ácido nucleico y el marcador.
La **Figura 15** representa un método de escisión de ácido nucleico diana multiplexado.
La **Figura 16** representa un método de administración estequiométrica de ácidos nucleicos de ARN.
La **Figura 17** representa un método de administración estequiométrica de ácidos nucleicos.
La **Figura 18** representa una inserción perfecta de un elemento indicador en un ácido nucleico diana usando un polipéptido dirigido a un sitio de la divulgación.
La **Figura 19** representa la eliminación de un elemento indicador de un ácido nucleico diana.
La **Figura 20** representa partes complementarias de las secuencias de ácido nucleico de un ácido nucleico pre-CRISPR y secuencias de ácido nucleico crtra de *Streptococcus pyogenes* SF370.
La **Figura 21** representa una estructura secundaria ilustrativa de un ácido nucleico dirigido a ácido nucleico guía en una sola secuencia sintético.
Las **Figuras 22A y B** muestran ejemplos de variantes de cadena principal de ácido nucleico dirigido a ácido nucleico guía sencillo. Los nucleótidos de los recuadros corresponden a nucleótidos que han sido alterados en relación con las secuencias de CRISPR marcadas como secuencia de ARNcr-crtra de longitud completa.
La **Figura 23A-C** muestra datos ilustrativos de un ensayo de escisión *in vitro*. Los resultados demuestran que más de una secuencia sintética de cadena principal de ácido nucleico dirigido a ácido nucleico puede soportar la escisión por un polipéptido dirigido a un sitio (por ejemplo, Cas9).
La **Figura 24** muestra ejemplos de secuencias sintéticas de ácido nucleico dirigido a ácido nucleico guía sencillo que contienen variantes en la región/el dúplex complementario. Los nucleótidos de los recuadros corresponden a nucleótidos que han sido alterados en relación con las secuencias de CRISPR marcadas como secuencia de ARNcr-crtra de longitud completa.
La **Figura 25** muestra variantes ilustrativas de la estructura de ácido nucleico dirigido a ácido nucleico guía sencillo dentro de la región 3' con respecto a la región/al dúplex complementario. Los nucleótidos de los recuadros corresponden a nucleótidos que han sido alterados en relación con el emparejamiento de secuencias de ácido nucleico crtra y ácido nucleico de CRISPR de *S. pyogenes* SF370 natural.
La **Figura 26A-B** muestra variantes ilustrativas de la estructura de ácido nucleico dirigido a ácido nucleico guía sencillo dentro de la región 3' con respecto a la región/al dúplex complementario. Los nucleótidos de los recuadros corresponden a nucleótidos que han sido alterados en relación con el emparejamiento de secuencias de ácido nucleico crtra y ácido nucleico de CRISPR de *S. pyogenes* SF370 natural.
La **Figura 27A-B** muestra variantes de estructuras de ácido nucleico dirigido a ácido nucleico ilustrativas que comprenden secuencias de horquilla adicionales derivadas de la repetición CRISPR en *Pseudomonas aeruginosa* (PA 14). Las secuencias de los recuadros pueden unirse a la ribonucleasa Csy4 de PAH.
La **Figura 28** muestra datos ilustrativos de un ensayo de escisión *in vitro* que demuestra que múltiples secuencias de cadena principal sintéticas de ácido nucleico dirigido a ácido nucleico soportan la escisión de Cas9. Las imágenes de gel superior e inferior representan dos repeticiones independientes del ensayo.
La **Figura 29** muestra datos ilustrativos de un ensayo de escisión *in vitro* que demuestra que múltiples secuencias de cadena principal sintéticas de ácido nucleico dirigido a ácido nucleico soportan la escisión de Cas9. Las imágenes de gel superior e inferior representan dos repeticiones independientes del ensayo.
La **Figura 30** representa métodos ilustrativos de la divulgación para llevar un polinucleótido donante a un sitio de modificación en un ácido nucleico diana.
La **Figura 31** muestra un sistema para almacenar y compartir información electrónica.
La **Figura 32** representa dos nickasas que generan un corte terminal romo en un ácido nucleico diana. No se muestran los polipéptidos modificadores dirigidos a un sitio complejados con ácidos nucleicos dirigidos a ácido nucleico.
La **Figura 33** muestra el corte escalonado de un ácido nucleico diana usando dos nickasas y generando extremos pegajosos. No se muestran los polipéptidos modificadores dirigidos a un sitio complejados con ácidos nucleicos dirigidos a ácido nucleico.
La **Figura 34** representa el corte escalonado de un ácido nucleico diana usando dos nickasas y generando extremos pegajosos de tamaño medio. No se muestran los polipéptidos modificadores dirigidos a un sitio complejados con ácidos nucleicos dirigidos a ácido nucleico.
La **Figura 35** ilustra una alineación de secuencias de ortólogos de Cas9. Los aminoácidos con una "X" debajo de

ellos pueden considerarse similares. Los aminoácidos con una "Y" debajo de ellos pueden considerarse altamente conservados o idénticos en todas las secuencias. Los restos de aminoácidos sin una "X" o una "Y" pueden no conservarse.

La **Figura 36** muestra la funcionalidad de las variantes de ácido nucleico dirigido a ácido nucleico en la escisión de ácido nucleico diana. Las variantes probadas en la *Figura 36* corresponden a las variantes representadas en la **Figura 22**, **Figura 24,** y **Figura 25**.

La **Figura 37A-D** muestra ensayos de escisión *in vitro* usando variantes de ácidos nucleicos dirigidos a ácido nucleico.

La **Figura 38** representa secuencias de aminoácidos ilustrativas de Csy4 de *P. aeruginosa* de tipo silvestre.

La **Figura 39** representa secuencias de aminoácidos ilustrativas de una endorribonucleasa enzimáticamente inactiva (por ejemplo, Csy4).

La **Figura 40** representa secuencias de aminoácidos ilustrativas de Csy4 de *P. aeruginosa*.

La **Figura 41A-J** representa secuencias de aminoácidos ilustrativas de Cas6.

La **Figura 42A-C** representa secuencias de aminoácidos ilustrativas de Cas6.

Descripción detallada de la invención

La invención se define por las reivindicaciones adjuntas.

Definiciones

Como se usa en el presente documento, "marcador de afinidad" puede referirse a cualquier marcador de afinidad peptídico o un marcador de afinidad de ácido nucleico. el marcador de afinidad se refiere, en general, a una secuencia de proteína o de ácido nucleico que puede unirse a una molécula (por ejemplo, unirse por una molécula pequeña, proteína, enlace covalente). Un marcador de afinidad puede ser una secuencia no nativa. Un marcador de afinidad peptídico puede comprender un péptido. Un marcador de afinidad peptídico puede ser aquel que pueda formar parte de un sistema dividido (por ejemplo, dos fragmentos de péptido inactivos pueden combinarse en *trans* para formar un marcador de afinidad activo). Un marcador de afinidad de ácido nucleico puede comprender un ácido nucleico. Un marcador de afinidad de ácido nucleico puede ser una secuencia que puede unirse selectivamente a una secuencia de ácido nucleico conocida (por ejemplo, mediante hibridación). Un marcador de afinidad de ácido nucleico puede ser una secuencia que puede unirse selectivamente a una proteína. Un marcador de afinidad puede fusionarse con una proteína nativa. Un marcador de afinidad puede fusionarse con una secuencia de nucleótidos. Algunas veces, un, dos o una pluralidad de marcadores de afinidad pueden fusionarse con una proteína nativa o secuencia de nucleótidos. Se puede introducir un marcador de afinidad en un ácido nucleico dirigido a ácido nucleico usando métodos de transcripción *in vitro* o *in vivo*. Los marcadores de afinidad de ácido nucleico pueden incluir, por ejemplo, un marcador químico, una secuencia de unión de proteínas de unión a ARN, una secuencia de unión de proteínas de unión a ADN, una secuencia hibridable a un polinucleótido marcado por afinidad, un aptámero de ARN sintético o un aptámero de ADN sintético. Los ejemplos de marcadores de afinidad de ácido nucleico químicos pueden incluir, pero sin limitación, ribo-nucleotrifosfatos que contienen biotina, colorantes fluorescentes y digoxeginina. Los ejemplos de marcadores de afinidad de ácido nucleico que se unen a proteínas pueden incluir, pero sin limitación, la secuencia de unión a MS2, la secuencia de unión a U1A, secuencias de proteínas de unión de tallo-lazo, la secuencia de boxB, la secuencia de eIF4A, o cualquier secuencia reconocida por una proteína de unión a ARN. Los ejemplos de oligonucleótidos marcados por afinidad de ácido nucleico pueden incluir, pero sin limitación, oligonucleótidos biotinilados, oligonucleótidos de 2,4-dinitrofenilo, oligonucleótidos de fluoresceína y oligonucleótidos conjugados con amina primaria.

Un marcador de afinidad de ácido nucleico puede ser un aptámero de ARN. Los aptámeros pueden incluir, aptámeros que se unen a la teofilina, estreptavidina, dextrano B512, adenosina, guanosina, guanina/xantina, 7-metil-GTP, aptámeros de aminoácidos tales como los aptámeros que se unen a la arginina, citrulina, valina, triptófano, cianocobalamina, N-metilmesoporfirina IX, flavina, NAD, y aptámeros de antibióticos tales como los aptámeros que se unen a la tobramicina, neomicina, lividomicina, kanamicina, estreptomicina, viomicina y cloranfenicol.

Un marcador de afinidad de ácido nucleico puede comprender una secuencia de ARN que puede estar unida por un polipéptido dirigido a un sitio. El polipéptido dirigido a un sitio puede ser condicionalmente inactivo enzimáticamente. La secuencia de ARN puede comprender una secuencia que puede estar unida por un miembro de los sistemas de CRISPR de tipo I, tipo II y/o tipo III. La secuencia de ARN puede estar unida por una proteína miembro de la familia de RAMP. La secuencia de ARN puede estar unida por una proteína miembro de la familia de Cas6 (por ejemplo, Csy4, Cas6). La secuencia de ARN puede estar unida por una proteína miembro de la familia de Cas5 (por ejemplo, Cas5). Por ejemplo, Csy4 puede unirse a una secuencia de horquilla de ARN específica con alta afinidad (Kd ~50 pM) y puede escindir el en un sitio 3' con respecto a la horquilla. La proteína miembro de la familia de Cas5 o Cas6 puede unirse a una secuencia de ARN que comprenda al menos aproximadamente o como máximo aproximadamente un 30 %, 40 %, 50 %, 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95% o 100% de identidad de secuencia y/o similitud de secuencia con las siguientes secuencias de nucleótidos: 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3; 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3'; 5'-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3'; 5'-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3'; 5'-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3'; 5'-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3'; 5'-

GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3’; 5’-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3’; 5’-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3’. 5’-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3’; 5’-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3’; 5’-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3’; 5’-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3’; 5’-GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3’; 5’-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3’; 5’-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3’, 5’-GUCGCCCCCCACGCGGGGGCGUGGAUUGAAAC-3 5’-CCAGCCGCCUUCGGGCGGCUGUGUGUUGAAAC-3’; 5’-GUCGCACUCUACAUGAGUGCGUGGAUUGAAAU-3’; 5’-UGUCGCACCUUAUAUAGGUGCGUGGAUUGAAAU-3 y 5’-GUCGCGCCCCGCAUGGGGCGCGUGGAUUGAAA-3’,

un marcador de afinidad de ácido nucleico puede comprender una secuencia de ADN que puede unirse por un polipéptido dirigido a un sitio. El polipéptido dirigido a un sitio puede ser condicionalmente inactivo enzimáticamente. La secuencia de ADN puede comprender una secuencia que puede estar unida por un miembro del sistema de CRISPR de tipo I, tipo II y/o tipo III. La secuencia de ADN puede estar unida por una proteína Argonauta. La secuencia de ADN puede estar unida por una proteína que contenga un dominio de dedo de cinc, un dominio TALE o cualquier otro dominio de unión a ADN.

Un marcador de afinidad de ácido nucleico puede comprender una secuencia de ribozima. Las ribozimas adecuadas pueden incluir ARNr de 23S de peptidil transferasa, RnasaP, intrones del grupo I, intrones del grupo II, ribozima de ramificación de GIR1, Leadzima, ribozimas de horquilla, ribozimas de cabeza de martillo, ribozimas de HDV, ribozimas de CPEB3, ribozimas de VS, ribozima de glmS, ribozima de CoTC y ribozimas sintéticas.

Los marcadores de afinidad peptídicos pueden comprender marcadores que pueden usarse para el seguimiento o la purificación (por ejemplo, una proteína fluorescente, proteína fluorescente verde (GFP), YFP, RFP, CFP, mCherry, tdTomato, un marcador His, (por ejemplo, un marcador 6XHis), un marcador de hemaglutinina (HA), un marcador FLAG, un marcador Myc, un marcador GST, un marcador MBP y un marcador de proteína de unión a quitina, un marcador de calmodulina, un marcador V5, un marcador de unión a estreptavidina y similares).

Los marcadores de afinidad de ácido nucleico y péptido pueden comprender marcadores de molécula pequeña tales como biotina o digitoxina, marcadores fluorescentes, tales como, por ejemplo, fluoresceína, rodamina, colorantes de flúor Alexa, colorante de cianina3, colorante de cianina5.

Los marcadores de afinidad de ácido nucleico pueden ubicarse a 5' de un ácido nucleico (por ejemplo, un ácido nucleico dirigido a ácido nucleico). Los marcadores de afinidad de ácido nucleico pueden ubicarse a 3' de un ácido nucleico. Los marcadores de afinidad de ácido nucleico pueden ubicarse a 5' y 3' de un ácido nucleico. Los marcadores de afinidad de ácido nucleico pueden ubicarse dentro de un ácido nucleico. Los marcadores de afinidad de péptidos se pueden ubicar en el extremo N-terminal de una secuencia polipeptídica. Los marcadores de afinidad peptídicos pueden ubicarse en el extremo C-terminal de una secuencia polipeptídica. Los marcadores de afinidad peptídicos pueden ubicarse en el extremo N-terminal y C-terminal de una secuencia polipeptídica. Se puede fusionar una pluralidad de marcadores de afinidad con un ácido nucleico y/o una secuencia polipeptídica.

Como se usa en el presente documento, "Agente de captura" puede referirse, en general, a un agente que puede purificar un polipéptido y/o un ácido nucleico. Un agente de captura puede ser una molécula o un material biológicamente activo (por ejemplo, cualquier sustancia biológica que se encuentre en la naturaleza o sintética, e incluye, pero sin limitación, células, virus, partículas subcelulares, proteínas, incluyendo más específicamente anticuerpos, inmunoglobulinas, antígenos, lipoproteínas, glucoproteínas, péptidos, polipéptidos, complejos proteicos, complejos de (estrepto)avidina-biotina, ligandos, receptores o moléculas pequeñas, aptámeros, ácidos nucleicos, ADN, ARN, ácidos nucleicos peptídicos, oligosacáridos, polisacáridos, lipopolisacáridos, metabolitos celulares, haptenos, sustancias farmacológicamente activas, alcaloides, esteroides, vitaminas, aminoácidos y azúcares). En algunos aspectos, el agente de captura puede comprender un marcador de afinidad. En algunos aspectos, un agente de captura puede unirse preferentemente a un polipéptido o ácido nucleico diana de interés. Los agentes de captura pueden flotar libremente en una mezcla. Los agentes de captura pueden unirse a una partícula (por ejemplo, una perla, una microperla, una nanopartícula). Los agentes de captura pueden unirse a una superficie sólida o semisólida. En algunos casos, los agentes de captura están unidos irreversiblemente a un diana. En otros casos, los agentes de captura están unidos de forma reversible a un diana (por ejemplo, si se puede eluir un diana o mediante el uso de un compuesto químico tal como el imidizol).

Como se usa en el presente documento, "Cas5" puede referirse, en general, a un polipéptido con al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas5 ilustrativo de tipo silvestre (por ejemplo, Cas5 de *D. vulgaris* y/o cualquier secuencia representada en la **Figura 42**). Cas5 puede referirse, en general, a un polipéptido con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas5 ilustrativo de tipo silvestre (por ejemplo, un Cas5 de *D. vulgaris*). Cas5 puede referirse al tipo silvestre o a una forma modificada de la proteína Cas5 que puede comprender

un cambio de aminoácido tal como una eliminación, inserción, sustitución, variante, mutación, fusión, quimera, o cualquier combinación de las mismas.

Como se usa en el presente documento, "Cas6" puede referirse, en general, a un polipéptido con al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas6 ilustrativo de tipo silvestre (por ejemplo, Cas6 de *T. thermophilus* y/o cualquier secuencia representada en la **Figura 41**). Cas6 puede referirse, en general, a un polipéptido con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas6 ilustrativo de tipo silvestre (por ejemplo, de *T. thermophilus*). Cas6 puede referirse al tipo silvestre o a una forma modificada de la proteína Cas6 que puede comprender un cambio de aminoácido tal como una eliminación, inserción, sustitución, variante, mutación, fusión, quimera, o cualquier combinación de las mismas.

Como se usa en el presente documento, "Cas9" puede referirse, en general, a un polipéptido con al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas9 ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes* (SEQ ID NO: 8, SEQ ID NO: 1-256, SEQ ID NO: 795-1346). Cas9 puede referirse a un polipéptido con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Cas9 ilustrativo de tipo silvestre (por ejemplo, de *S. pyogenes*). Cas9 puede referirse al tipo silvestre o una forma modificada de la proteína Cas9 que puede comprender un cambio de aminoácido tal como una eliminación, inserción, sustitución, variante, mutación, fusión, quimera, o cualquier combinación de las mismas.

Como se usa en el presente documento, una "célula" puede referirse, en general, a una célula biológica. Una célula puede ser la unidad estructural, funcional y/o biológica de un organismo vivo. Una célula puede proceder de cualquier organismo que tenga una o más células. Algunos ejemplos no limitantes incluyen: una célula procariota, célula eucariota, una célula bacteriana, una célula de arquea, una célula de un organismo eucariota unicelular, una célula de protozoo, una célula de una planta (por ejemplo, células de cultivos vegetales, frutas, hortalizas, cereales, semilla de soja, maíz, maíz, trigo, semillas, tomates, arroz, yuca, caña de azúcar, calabaza, heno, patatas, algodón, cannabis, tabaco, plantas en flor, coníferas, gimnospermas, helechos, *Lycopodiopsida*, hornabeques, hepáticas, musgos), una célula de algas, (por ejemplo,, *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens C. Agardh,* y similares), algas marinas (por ejemplo, alga parda), 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, mosca de la fruta, cnidario, equinodermo, nematodo, etc.), una célula de un animal vertebrado (por ejemplo, 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, etc.), etcétera. A veces, una célula no procede de un organismo natural (por ejemplo, una célula puede estar hecha sintéticamente, a veces, se denomina célula artificial).

Una célula puede estar *in vitro*. Una célula puede estar *in vivo*. Una célula puede ser una célula aislada. Una célula puede ser una célula de dentro de un organismo. Una célula puede ser un organismo. Una célula puede ser una célula en un cultivo celular. Una célula puede ser una de una colección de células. Una célula puede ser una célula procariota o puede derivarse de una célula procariota. Una célula puede ser una célula bacteriana o puede derivarse de una célula bacteriana. Una célula puede ser una célula de arquea o derivada de una célula de arquea. Una célula puede ser una célula eucariota o puede derivarse de una célula eucariota. Una célula puede ser una célula vegetal o puede derivarse de una célula vegetal. Una célula puede ser una célula animal o puede derivarse de una célula animal. Una célula puede ser una célula de invertebrado o puede derivarse de una célula de invertebrado. Una célula puede ser una célula de vertebrado o puede derivarse de una célula de vertebrado. Una célula puede ser una célula de mamífero o puede derivarse de una célula de mamífero. Una célula puede ser una célula de roedor o puede derivarse de una célula de roedor. Una célula puede ser una célula humana o puede derivarse de una célula humana. Una célula puede ser una célula microbiana o puede derivarse de una célula microbiana. Una célula puede ser una célula de hongo o puede derivarse de una célula de hongo.

Una célula puede ser una célula madre o una célula progenitora. Las células pueden incluir células madre (por ejemplo, células madre adultas, células iPS) y células progenitoras (por ejemplo, células progenitoras cardíacas, células progenitoras neuronales, etc.). Las células pueden incluir células madre y células progenitoras de mamíferos, incluyendo células madre de roedores, células progenitoras de roedores, células madre humanas, células progenitoras humanas, etc. Las células clónicas pueden comprender la progenie de una célula. Una célula puede comprender un ácido nucleico diana. Una célula puede estar en un organismo vivo. Una célula puede ser una célula modificada genéticamente. Una célula puede ser una célula hospedadora.

Una célula puede ser una célula madre totipotente, sin embargo, en algunas realizaciones de la divulgación, se puede usar el término "célula", pero puede no referirse a una célula madre totipotente. Una célula puede ser una célula vegetal, pero, en algunas realizaciones de la divulgación, se puede usar el término "célula", pero puede no referirse a una célula vegetal. Una célula puede ser una célula pluripotente. Por ejemplo, una célula puede ser una célula hematopoyética pluripotente que puede diferenciarse en otras células en el linaje celular hematopoyético, pero puede

no ser capaz de diferenciarse en otra célula no hematopoyética. Una célula puede convertirse en un organismo completo. Una célula puede o no ser capaz de convertirse en un organismo completo. Una célula puede ser un organismo completo.

Una célula puede ser una célula primaria. Por ejemplo, los cultivos de células primarias se pueden pasar 0 veces, 1 vez, 2 veces, 4 veces, 5 veces, 10 veces, 15 veces o más. Las células pueden ser organismos unicelulares. Las células pueden crecer en cultivo.

Una célula puede ser una célula enferma. Una célula enferma puede tener alteración del metabolismo, de la expresión génica y/o de las características morfológicas. Una célula enferma puede ser una célula cancerosa, una célula diabética y una célula apoptótica. Una célula enferma puede ser una célula de un sujeto enfermo. Las enfermedades ilustrativas pueden incluir trastornos de la sangre, cánceres, trastornos metabólicos, trastornos oculares, trastornos de órganos, trastornos musculoesqueléticos, enfermedad cardíaca y similares.

Si las células son células primarias, pueden cosecharse de un individuo por cualquier método. Por ejemplo, los leucocitos pueden cosecharse por aféresis, leucocitaféresis, separación por gradiente de densidad, etc. Las células de tejidos tales como la piel, músculo, médula ósea, bazo, hígado, páncreas, pulmón, intestino, estómago, etc. se pueden cosechar mediante biopsia. Puede utilizarse una solución apropiada para la dispersión o suspensión de las células cosechadas. Dicha solución puede ser, en general, una solución de sal equilibrada, (por ejemplo, solución salina normal, solución salina tamponada con fosfato (PBS), solución de sal equilibrada de Hank, etc.), convenientemente complementada con suero de ternera fetal u otros factores de origen natural, junto con un tampón aceptable a baja concentración. Los tampones pueden incluir HEPES, tampones de fosfato, tampones de lactato, etc. Las células se pueden usar de inmediato o se pueden almacenar (por ejemplo, por congelación). Las células congeladas pueden descongelarse y pueden reutilizarse. Las células pueden congelarse en DMSO, suero, tampón de medio (por ejemplo, DMSO al 10 %, suero al 50 %, medio tamponado al 40 %), y/o alguna otra solución común usada para preservar las células a temperaturas de congelación.

Como se usa en el presente documento, "polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente", en general, puede referirse a un polipéptido que puede unirse a una secuencia de ácido nucleico en un polinucleótido de una manera específica de la secuencia, pero no puede escindir un polinucleótido diana, excepto en una o más condiciones que activen el dominio enzimático. Un polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente puede comprender un dominio inactivo enzimáticamente que puede activarse condicionalmente. Un polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente puede activarse condicionalmente en presencia de imidazol. Un polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente puede comprender un sitio activo mutado que no se una a su ligando afín, dando lugar a un polipéptido dirigido a un sitio inactivo enzimáticamente. El sitio activo mutado puede diseñarse para unirse a análogos de ligando, de modo que un análogo de ligando puede unirse al sitio activo mutado y reactivo al polipéptido dirigido a un sitio. Por ejemplo, las proteínas de unión a ATP pueden comprender un sitio activo mutado que puede inhibir la actividad de la proteína, sin embargo, están diseñadas para unirse específicamente a los análogos de ATP. La unión de un análogo de ATP, pero no de ATP, puede reactivar la proteína. El polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente puede comprender una o más secuencias no nativas (por ejemplo, una fusión, un marcador de afinidad).

Como se usa en el presente documento, "ARNcr" puede referirse, en general, a un ácido nucleico con al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un ARNcr ilustrativo de tipo silvestre (por ejemplo, un ARNcr de *S. pyogenes* (por ejemplo, SEQ ID NO: 569, SEQ ID NO: 563-679). El ARNcr puede referirse, en general, a un ácido nucleico con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100% de identidad de secuencia y/o similitud de secuencia con un ARNcr ilustrativo de tipo silvestre (por ejemplo, un ARNcr de *S. pyogenes*). El ARNcr puede referirse a una forma modificada de un ARNcr que puede comprender un cambio de nucleótido tal como una eliminación, inserción o sustitución, variante, mutación o quimera. Un ARNcr puede ser un ácido nucleico que sea al menos aproximadamente un 60 % idéntico a una secuencia de ARNcr ilustrativa de tipo silvestre (por ejemplo, un ARNcr de *S. pyogenes*) sobre un tramo de al menos 6 nucleótidos contiguos. Por ejemplo, una secuencia de ARNcr puede ser al menos aproximadamente un 60 % idéntica, al menos aproximadamente un 65 % idéntica, al menos aproximadamente un 70 % idéntica, al menos aproximadamente un 75 % idéntica, al menos aproximadamente un 80 % idéntica, al menos aproximadamente un 85 % idéntica, al menos aproximadamente un 90 % idéntico, al menos aproximadamente un 95 % idéntica, al menos aproximadamente un 98 % idéntica, al menos aproximadamente un 99 % idéntica o un 100 % idéntica, a una secuencia de ARNcr ilustrativa de tipo silvestre (por ejemplo, un ARNcr de *S. pyogenes*) en un tramo de al menos 6 nucleótidos contiguos.

Como se usa en el presente documento, "Repetición CRISPR" o "Secuencia de repetición CRISPR" puede referirse a una secuencia de repetición CRISPR mínima.

Como se usa en el presente documento, "Cas4" puede referirse, en general, a un polipéptido con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Csy4 ilustrativo de tipo silvestre (por ejemplo, Csy4 de *P. aeruginosa*, véase la **Figura 40**). "Cas4" puede referirse, en general, a un polipéptido con al menos aproximadamente un 5 %,

10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con un polipéptido Csy4 ilustrativo de tipo silvestre (por ejemplo, Csy4 de *P. aeruginosa*). Csy4 puede referirse al tipo silvestre o una forma modificada de la proteína Csy4 que puede comprender un cambio de aminoácido tal como una eliminación, inserción, sustitución, variante, mutación, fusión, quimera, o cualquier combinación de las mismas.

Como se usa en el presente documento, La "endorribonucleasa" puede referirse, en general, a un polipéptido que puede escindir ARN. En algunos aspectos, una endorribonucleasa puede ser un polipéptido dirigido a un sitio. Una endorribonucleasa puede ser miembro de un sistema de CRISPR (por ejemplo, Tipo I, Tipo II, Tipo III). La endorribonucleasa puede referirse a una superfamilia de proteínas misteriosas asociadas a repeticiones (RAMP) (por ejemplo, Cas6, familias Cas6, Cas5). Las endorribonucleasas también pueden incluir los miembros de las familias RNasa A, RNasa H, RNasa I, RNasa III (por ejemplo, Drosha, Dicer, RNasa N), RNasa L, RNasa P, RNasa PhyM, RNasa Tl, RNasa T2, RNasa U2, RNasa V1, RNasa V. Una endorribonucleasa puede referirse a una endorribonucleasa condicionalmente enzimáticamente inactiva. Una endorribonucleasa puede referirse a una endorribonucleasa catalíticamente inactiva.

Como se usa en el presente documento, "polinucleótido donante" puede referirse a un ácido nucleico que puede integrarse en un sitio durante el diseño del genoma o el diseño del ácido nucleico diana.

Como se usa en el presente documento, "fijador" o "agente de entrecruzamiento" puede referirse, en general, a un agente que puede fijar o entrecruzar células. Las células fijas o entrecruzadas pueden estabilizar los complejos de proteína-ácido nucleico en la célula. Los fijadores y agentes de entrecruzamiento adecuados pueden incluir, formaldehído, glutaraldehído, fijadores a base de etanol, fijadores a base de metanol, acetona, ácido acético, tetraóxido de osmio, dicromato de potasio, ácido crómico, permanganato de potasio, mercuriales, picratos, formol, paraformaldehído, agentes de entrecruzamiento de NHS-éster reactivos con amina tales como bis[sulfosuccinimidil]suberato (BS3), 3,3'-ditiobis[sulfosuccinimidilpropionato] (DTSSP), bis[sulfosuccinimidilsuccinato de etilenglicol (sulfo-EGS), glutarato de disuccinimidilo (DSG), propionato de ditiobis[succinimidilo] (DSP), suberato de disuccinimidilo (DSS), bis[succinimidilsuccinato] de etilenglicol (EGS), agentes de entrecruzamiento de NHS-éster/diazirina tales como NHS-diazirina, NHS-LC-diazirina, NHS-SS-diazirina, sulfo-NHS-diazirina, sulfo-NHS-LC-diazirina y sulfo-NHS-SS-diazirina.

Como se usa en el presente documento, "Fusión" puede referirse a una proteína y/o a un ácido nucleico que comprende una o más secuencias no nativas (por ejemplo, fracciones). Una fusión puede comprender una o más secuencias no nativas iguales. Una fusión puede comprender una o más secuencias no nativas diferentes. Una fusión puede ser una quimera. Una fusión puede comprender un marcador de afinidad de ácido nucleico. Una fusión puede comprender un código de barras. Una fusión puede comprender un marcador de afinidad peptídico. Una fusión puede proporcionar la localización subcelular del polipéptido dirigido a un sitio (por ejemplo, una señal de localización nuclear (NLS) para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a las mitocondrias, una señal de localización de cloroplasto para dirigirse a un cloroplasto, una señal de retención del retículo endoplasmático (RE) y similares). Una fusión puede proporcionar una secuencia no nativa (por ejemplo, marcador de afinidad) que se puede usar para rastrear o purificar. Una fusión puede ser una molécula pequeña tal como biotina o un colorante tal como los colorantes de flúor alexa, colorante de cianina3, colorante de cianina5. La fusión puede proporcionar una mayor o menor estabilidad.

En algunas realizaciones, una fusión puede comprender un marcador detectable, incluyendo una fracción que puede proporcionar una señal detectable. Los marcadores y/o fracciones detectables adecuados que pueden proporcionar una señal detectable pueden incluir, pero sin limitación, una enzima, un radioisótopo, un miembro de un par de unión específico; un fluoróforo; una proteína fluorescente; un punto cuántico; y similares.

Una fusión puede comprender un miembro de un par de FRET. Los pares de FRET (donante/aceptor) adecuados para su uso pueden incluir, pero sin limitación, EDANS/fluoresceína, IAEDANS/fluoresceína, fluoresceína/tetrametilrodamina, fluoresceína/Cy 5, IEDANS/DABCYL, fluoresceína/QSY-7, fluoresceína/lC Rojo 640, fluoresceína/Cy 5.5 y fluoresceína/lC Rojo 705.

Un par de donante/aceptor de fluoróforo/punto cuántico puede usarse como fusión. Los fluoróforos adecuados ("marcador fluorescente") pueden incluir cualquier molécula que pueda detectarse a través de sus propiedades fluorescentes inherentes, que puede incluir fluorescencia detectable tras la excitación. Los marcadores fluorescentes adecuados pueden incluir, pero sin limitación, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metil-cumarinas, pireno, Verde de malacita, estilbeno, Lucifer Yellow, Cascade Blue™, Texas Red, IAEDANS, EDANS, BODIPY FL, LC Rojo 640, Cy 5, Cy 5.5, LC Rojo 705 y Oregon green.

Una fusión puede comprender una enzima. Las enzimas adecuadas pueden incluir, pero sin limitación, peroxidasa de rábano picante, luciferasa, beta-galactosidasa y similares.

Una fusión puede comprender una proteína fluorescente. Las proteínas fluorescentes adecuadas pueden incluir, pero sin limitación, una proteína fluorescente verde (GFP), (por ejemplo, una GFP de *Aequoria victoria,* proteínas fluorescentes de *Anguilla japonica*, o un mutante o derivado de las mismas), una proteína fluorescente roja, una

proteína fluorescente amarilla, cualquiera de varias proteínas fluorescentes y coloreadas.

Una fusión puede comprender una nanopartícula. Las nanopartículas adecuadas pueden incluir nanopartículas fluorescentes o luminiscentes y nanopartículas magnéticas. Se puede detectar cualquier propiedad o característica óptica o magnética de la/s nanopartícula/s.

Una fusión puede comprender puntos cuánticos (QD). Los QD pueden volverse solubles en agua aplicando capas de recubrimiento que comprendan varios materiales diferentes. Por ejemplo, los QD se pueden solubilizar usando polímeros anfífilos. Los polímeros ilustrativos que se han empleado pueden incluir ácido poliacrílico de bajo peso molecular modificado con octilamina, fosfolípidos derivados de polietilenglicol (PEG), polianhídridos, copolímeros de bloque, etc. Los QD se pueden conjugar con un polipéptido a través de cualquiera de varios grupos funcionales o agentes de enlace diferentes que se pueden unir directa o indirectamente a una capa de recubrimiento. Hay QD con una amplia variedad de espectros de absorción y emisión disponibles en el mercado, por ejemplo, de Quantum Dot Corp. (Hayward Calif.; ahora propiedad de Invitrogen) o de Evident Technologies (Troy, N.Y.). Por ejemplo, Hay QD que tienen longitudes de onda de emisión máxima de aproximadamente 525, 535, 545, 565, 585, 605, 655, 705 y 800 nm. Por lo tanto, los QD pueden tener un intervalo de colores diferentes en la parte visible del espectro y, en algunos casos, incluso más allá.

Los radioisótopos adecuados pueden incluir, pero sin limitación, ^{14}C, ^{3}H, ^{32}P, ^{33}P, ^{35}S y ^{125}I.

Como se usa en el presente documento, "célula modificada genéticamente" puede referirse, en general, a una célula que ha sido modificada genéticamente. Algunos ejemplos no limitantes de modificaciones genéticas pueden incluir: inserciones, eliminaciones, inversiones, translocaciones, fusiones de genes o cambio de uno o más nucleótidos. Una célula modificada genéticamente puede comprender un ácido nucleico diana con una rotura de doble cadena introducida (por ejemplo, rotura de ADN). Una célula modificada genéticamente puede comprender un ácido nucleico introducido exógenamente (por ejemplo, un vector). Una célula modificada genéticamente puede comprender un polipéptido introducido exógenamente de la divulgación y/o ácido nucleico de la divulgación. Una célula modificada genéticamente puede comprender un polinucleótido donante. Una célula modificada genéticamente puede comprender un ácido nucleico exógeno integrado en el genoma de la célula modificada genéticamente. Una célula modificada genéticamente puede comprender una eliminación de ADN. Una célula modificada genéticamente también puede referirse a una célula con ADN mitocondrial o de cloroplasto modificado.

Como se usa en el presente documento, el "diseño del genoma" puede referirse a un proceso de modificación de un ácido nucleico diana. El diseño del genoma puede referirse a la integración de ácido nucleico no nativo en ácido nucleico nativo. El diseño del genoma puede referirse a la dirección de un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico a un ácido nucleico diana, sin una integración o una eliminación del ácido nucleico diana. El diseño del genoma puede referirse a la escisión de un ácido nucleico diana y la unión de nuevo del ácido nucleico diana sin una integración de una secuencia exógena en el ácido nucleico diana o una eliminación en el ácido nucleico diana. El ácido nucleico nativo puede comprender un gen. El ácido nucleico no nativo puede comprender un polinucleótido donante. En los métodos de la divulgación, los polipéptidos dirigidos a un sitio (por ejemplo, Cas9) pueden introducir roturas bicatenarias en ácido nucleico, (por ejemplo, ADN genómico). La rotura de doble cadena puede estimular las vías de reparación de ADN endógeno de una célula (por ejemplo, recombinación homóloga (HR) y/o la unión de extremos no homólogos (NHEJ) o A-NHEJ (unión de extremos no homólogos alternativa)). Se pueden introducir mutaciones, eliminaciones, alteraciones e integraciones de ácido nucleico foráneo, exógeno y/o alternativo en el sitio de la rotura del ADN bicatenario.

Como se usa en el presente documento, el término "aislado" puede referirse a un ácido nucleico o polipéptido que, de la mano de un ser humano, existe aparte de su entorno nativo y, por lo tanto, no es un producto de la naturaleza. Aislado puede significar esencialmente puro. Un ácido nucleico o polipéptido aislado puede existir en una forma purificada y/o puede existir en un entorno no nativo tal como, por ejemplo, en una célula transgénica.

Como se usa en el presente documento, "no nativo" puede referirse a una secuencia de ácido nucleico o polipéptido que no se encuentra en un ácido nucleico o proteína nativo. No nativo puede referirse a marcadores de afinidad. No nativo puede referirse a fusiones. No nativo puede referirse a una secuencia de ácido nucleico o polipéptido natural que comprende mutaciones, inserciones y/o eliminaciones. Una secuencia no nativa puede presentar y/o codificar una actividad (por ejemplo, actividad enzimática, actividad de metiltransferasa, actividad de acetiltransferasa, actividad de quinasa, actividad ubiquitinante, etc.) que también puede ser presentada por la secuencia de ácido nucleico y/o polipéptido a la que se fusiona la secuencia no nativa. Una secuencia de ácido nucleico o polipéptido no nativa puede unirse a una secuencia de ácido nucleico o polipéptido natural (o una variante de la misma) mediante ingeniería genética para generar una secuencia de ácido nucleico y/o polipéptido quimérica que codifique un ácido nucleico y/o polipéptido quimérico. Una secuencia no nativa puede referirse a una secuencia de extensión de la hibridación 3'.

Como se usa en el presente documento, un "ácido nucleico" puede referirse, en general, a una secuencia polinucleotídica o a un fragmento de la misma. Un ácido nucleico puede comprender nucleótidos. Un ácido nucleico puede ser exógeno o endógeno a una célula. Un ácido nucleico puede existir en un entorno libre de células. Un ácido nucleico puede ser un gen o un fragmento del mismo. Un ácido nucleico puede ser ADN. Un ácido nucleico puede ser

ARN. Un ácido nucleico puede comprender uno o más análogos (por ejemplo, cadena principal alterada, azúcar o nucleobase). Algunos ejemplos no limitantes de análogos incluyen: 5-bromouracilo, ácido nucleico peptídico, ácido xenonucleico, morfolinos, ácidos nucleicos bloqueados, ácidos nucleicos de glicol, ácidos nucleicos de treosa, didesoxinucleótidos, cordicepina, 7-desaza-GTP, fluoróforos (por ejemplo, rodamina o fluoresceína unidos al azúcar), nucleótidos que contienen tiol, nucleótidos unidos a biotina, análogos de bases fluorescentes, islas CpG, metil-7-guanosina, nucleótidos metilados, inosina, tiouridina, seudourdina, dihidrouridina, queuosina y wiosina.

Como se usa en el presente documento, una "muestra de ácido nucleico" puede referirse, en general, a una muestra de una entidad biológica. Una muestra de ácido nucleico puede comprender ácido nucleico. El ácido nucleico de la muestra de ácido nucleico estar purificado y/o enriquecido. La muestra de ácido nucleico puede mostrar la naturaleza de conjunto. Las muestras de ácido nucleico pueden proceder de varias fuentes. Las muestras de ácido nucleico pueden proceder de uno o más individuos. Una o más muestras de ácido nucleico pueden proceder del mismo individuo. Un ejemplo no limitante sería si una muestra procediera de la sangre de un individuo y una segunda muestra procediera de la biopsia tumoral de un individuo. Los ejemplos de muestras de ácido nucleico pueden incluir, entre otros, sangre, suero, plasma, hisopo nasal o hisopo nasofaringeo, saliva, orina, fluido gástrico, líquido espinal, lágrimas, deposición, mucosidad, sudor, cerumen, aceite, secreción glandular, líquido cefalorraquídeo, tejido, semen, fluido vaginal, líquidos intersticiales, incluidos los líquidos intersticiales derivados del tejido tumoral, líquidos oculares, líquido espinal, hisopo de garganta, hisopo de mejilla, aire espirado, cabello, uñas, piel, biopsia, líquido placentario, fluido amniótico, sangre de cordón umbilical, fluidos enfáticos, fluidos de cavidades, esputo, pus, micropiota, meconio, leche materna, muestras bucales, lavado nasofaríngeo, otras excreciones o cualquier combinación de las mismas. Las muestras de ácido nucleico pueden proceder de tejidos. Los ejemplos de muestras de tejido pueden incluir, entre otros, tejido conjuntivo, tejido muscular, tejido nervioso, tejido epitelial, cartílago, muestra cancerosa o tumoral, médula ósea o hueso. La muestra de ácido nucleico puede proporcionarse de un ser humano o animal. La muestra de ácido nucleico puede proporcionarse de un mamífero, vertebrado, tales como murinos, de simio, de seres humanos, de animales de granja, animales deportivos o animales domésticos. La muestra de ácido nucleico puede recogerse de un sujeto vivo o muerto. La muestra de ácido nucleico puede ser recién recogida de un sujeto o puede haber sido sometida a alguna forma de procesamiento previo, almacenamiento o transporte.

Una muestra de ácido nucleico puede comprender un ácido nucleico diana. Una muestra de ácido nucleico puede proceder de lisado celular. El lisado celular puede proceder de una célula.

Como se usa en el presente documento, "ácido nucleico dirigido a ácido nucleico" puede referirse a un ácido nucleico que puede hibridarse con otro ácido nucleico. Un ácido nucleico dirigido a ácido nucleico puede ser ARN. Un ácido nucleico dirigido a ácido nucleico puede ser ADN. El ácido nucleico dirigido a ácido nucleico puede programarse para unirse a una secuencia de ácido nucleico específica de un sitio. El ácido nucleico que se va a dirigido o el ácido nucleico diana, puede comprender nucleótidos. El ácido nucleico dirigido a ácido nucleico puede comprender nucleótidos. Una parte del ácido nucleico diana puede ser complementaria a una parte del ácido nucleico dirigido a ácido nucleico. Un ácido nucleico dirigido a ácido nucleico puede comprender una cadena polinucleotídica y puede denominarse un "ácido nucleico guía sencillo" (es decir, un "ácido nucleico dirigido a ácido nucleico guía sencillo"). Un ácido nucleico dirigido a ácido nucleico puede comprender dos cadenas polipeptídicas y puede denominarse un "ácido nucleico guía doble" (es decir, un "ácido nucleico dirigido a ácido nucleico guía doble"). Si no se especifica lo contrario, la expresión "ácido nucleico dirigido a ácido nucleico" puede ser inclusiva, refiriéndose tanto a los ácidos nucleicos guía sencillos como a los ácidos nucleicos guía dobles.

Un ácido nucleico dirigido a ácido nucleico puede comprender un segmento que puede denominarse un "segmento dirigido a ácido nucleico" o una "secuencia dirigida al ácido nucleico". Un ácido nucleico dirigido a ácido nucleico puede comprender un segmento que puede denominarse "segmento de unión a proteínas" o "secuencia de unión a proteínas".

Un ácido nucleico dirigido a ácido nucleico puede comprender una o más modificaciones (por ejemplo, una modificación de base, una modificación de la estructura principal), para proporcionar al ácido nucleico una característica nueva o mejorada (por ejemplo, estabilidad mejorada). Un ácido nucleico dirigido a ácido nucleico puede comprender un marcador de afinidad de ácido nucleico. Un nucleósido puede ser una combinación de base y azúcar. La parte de base del nucleósido puede ser una base heterocíclica. Las dos clases más comunes de dichas bases heterocíclicas son las purinas y las pirimidinas. Los nucleótidos pueden ser nucleósidos que además incluyen un grupo fosfato unido covalentemente a la parte de azúcar del nucleósido. Para aquellos nucleósidos que incluyen un azúcar pentofuranosilo, el grupo fosfato se puede unir a la fracción hidroxilo 2', 3' o 5' del azúcar. En la formación de ácidos nucleicos dirigidos a ácido nucleico, los grupos fosfato pueden unir covalentemente nucleósidos adyacentes entre sí para formar un compuesto polimérico lineal. A su vez, se pueden unir los respectivos extremos de este compuesto polimérico lineal más para formar un compuesto circular; sin embargo, en general, son adecuados los compuestos lineales. Además, los compuestos lineales pueden tener una complementariedad interna de las bases de nucleótidos y, por lo tanto, pueden plegarse de manera que se produzca un compuesto total o parcialmente bicatenario. Dentro de los ácidos nucleicos dirigidos a ácido nucleico, los grupos fosfato pueden denominarse comúnmente como la formación de la cadena principal internucleosídica del ácido nucleico dirigido a ácido nucleico. El enlace o la cadena principal del ácido nucleico dirigido a ácido nucleico puede ser un enlace fosfodiéster de 3' a 5'.

Un ácido nucleico dirigido a ácido nucleico puede comprender una cadena principal modificada y/o enlaces internucleosídicos modificados. Las cadenas principales modificadas pueden incluir aquellas que conservan un átomo de fósforo en la cadena principal y aquellas que no tienen un átomo de fósforo en la cadena principal.

Las cadenas principales de ácido nucleico dirigido a ácido nucleico modificadas adecuadas que contienen un átomo de fósforo pueden incluir, por ejemplo, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, metil- y otros alquil-fosfonatos tales como los 3'-alquilen-fosfonatos, 5'-alquilen-fosfonatos, fosfonatos quirales, fosfinatos, fosforamidatos que incluyen 3'-amino-fosforamidato y amino-alquilfosforamidatos, fosforodiamidatos, tionofosforamidatos, tionoalquilfosfonatos, tionoalquilfosfotriésteres, selenofosfatos y boranofosfatos que tienen enlaces 3-5' normales, análogos unidos en 2-5', y aquellos que tienen polaridad invertida, en donde uno o más enlaces internucleotídicos es un enlace de 3' a 3', de 5 'a 5' o de 2 'a 2'. Los ácidos nucleicos dirigidos a ácido nucleico adecuados que tienen polaridad invertida pueden comprender un solo enlace de 3' a 3' en el enlace internucleotídico más a 3' (es decir, un solo resto de nucleósido invertido en donde falta la nucleobase o tiene un grupo hidroxilo en su lugar). También se pueden incluir diversas sales (por ejemplo, cloruro de potasio o cloruro de sodio), sales mixtas y formas de ácido libre.

Un ácido nucleico dirigido a ácido nucleico puede comprender uno o más enlaces internucleosídicos de fosforotioato y/o heteroátomo, en particular $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (es decir, una cadena principal de metileno (metilimino) o MMI), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ y $-O-N(CH_3)-CH_2-CH_2-$ (en donde el enlace internucleotídico de fosfodiéster nativo se representa como $-O-P(=O)(OH)-O-CH_2-$).

Un ácido nucleico dirigido a ácido nucleico puede comprender una estructura de cadena principal de morfolino. Por ejemplo, un ácido nucleico puede comprender un anillo de morfolino de 6 elementos en lugar de un anillo de ribosa. En algunas de estas realizaciones, un enlace fosforodiamidato u otro enlace internucleosídico no fosfodiéster puede reemplazar a un enlace fosfodiéster.

Un ácido nucleico dirigido a ácido nucleico puede comprender cadenas principales polinucleotídicas que están formadas por enlaces internucleosídicos de alquilo o cicloalquilo de cadena corta, enlaces internucleosídicos de heteroátomos mixtos y de alquilo o cicloalquilo, o uno o más enlaces internucleosídicos heteroatómicos o heterocíclicos de cadena corta. Estos pueden incluir aquellos que tienen enlaces morfolino (formados en parte a partir de la parte de azúcar de un nucleósido); cadenas principales de siloxano; cadenas principales de sulfuro, sulfóxido y sulfona; cadenas principales de formacetilo y tioformacetilo; cadenas principales de metilen formacetilo y tioformacetilo; cadenas principales de riboacetilo; cadenas principales que contienen alqueno; cadenas principales de sulfamato; cadenas principales de metilenimino y metilenhidrazino; cadenas principales de sulfonato y sulfonamida; cadenas principales de amida; y otros que tienen partes componentes de N, O, S y CH2.

Un ácido nucleico dirigido a ácido nucleico puede comprender un mimético de ácido nucleico. El término "mimético" puede incluir polinucleótidos en los que solo el anillo de furanosa o tanto el anillo de furanosa como el enlace internucleotídico se reemplazan por grupos no furanosa, el reemplazo de solo el anillo de furanosa también puede denominarse sustituto del azúcar. Se puede mantener la fracción de base heterocíclica o una fracción de base heterocíclica modificada para la hibridación con un ácido nucleico diana apropiado. Uno de dichos ácidos nucleicos puede ser un ácido nucleico peptídico (ANP). En un ANP, la cadena principal de azúcar de un polinucleótido puede reemplazarse por una cadena principal que contenga amida, en particular, una cadena principal de aminoetilglicina. Los nucleótidos pueden ser conservados y están unidos directa o indirectamente a átomos de nitrógeno aza de la parte amida de la cadena principal. La cadena principal de los compuestos de ANP puede comprender dos o más unidades de aminoetilglicina unidas que dan al ANP una cadena principal que contiene una amida. Las fracciones de base heterocíclica pueden unirse directa o indirectamente a átomos de nitrógeno aza de la parte amida de la cadena principal.

Un ácido nucleico dirigido a ácido nucleico puede comprender unidades morfolino enlazadas (es decir, ácido nucleico de morfolino) que tienen bases heterocíclicas unidas al anillo morfolino. Los grupos de enlace pueden unir las unidades monoméricas de morfolino en un ácido nucleico de morfolino. Los compuestos oligoméricos a base de morfolino no iónicos pueden tener interacciones menos indeseables con las proteínas celulares. Los polinucleótidos a base de morfolino pueden ser miméticos no iónicos de ácidos nucleicos dirigidos a ácidos nucleicos. Se puede unir varios compuestos dentro de la clase morfolino usando diferentes grupos de enlace. Una clase adicional de polinucleótido mimético se puede denominar ácidos nucleicos de ciclohexenilo (CeNA). El anillo de furanosa normalmente presente en una molécula de ácido nucleico puede reemplazarse por un anillo de ciclohexenilo. Los monómeros de fosforamidita protegidos con DMT CeNA pueden prepararse y usarse para la síntesis de compuestos oligoméricos usando química de fosforamidita. La incorporación de monómeros de CeNA a una cadena de ácido nucleico puede aumentar la estabilidad de un híbrido de ADN/ARN. Los oligoadenilatos de CeNA pueden formar complejos con complementos de ácido nucleico con una estabilidad similar a los complejos nativos. Una modificación adicional puede incluir ácidos nucleicos bloqueados (LNA) en los que el grupo 2'-hidroxilo está unido al átomo de carbono 4' del anillo de azúcar, formando así un enlace 2'-C,4'-C-oximetileno, que da lugar a una fracción de azúcar bicíclica. El enlace puede ser un metileno ($-CH_2-$), grupo que une el átomo de oxígeno 2' y el átomo de carbono 4', en donde n es 1 o 2. Los LNA y análogos de LNA pueden mostrar estabilidades térmicas de dúplex muy altas con ácido nucleico complementario (Tf = +3 a +10 ºC), estabilidad hacia la degradación exonucleolítica en 3' y buenas propiedades de solubilidad.

Un ácido nucleico dirigido a ácido nucleico puede comprender una o más fracciones de azúcar sustituidas. Los polinucleótidos adecuados pueden comprender un grupo sustituyente de azúcar seleccionado de: OH; F; O-, S- o N-alquilo; O-, S- o N-alquenilo; O-, S- o N-alquinilo; u O-alquil-O-alquilo, en donde el alquilo, alquenilo y alquinilo pueden ser alquilo C_1 a C_{10} sustituido o no sustituido, o alquenilo y alquinilo C_2 a C_{10}. Son particularmente adecuados $O((CH_2)nO)\ mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$ y $O(CH_2)_nON((CH_2)nCH_3)_2$, en donde n y m son de 1 a aproximadamente 10. Se puede seleccionar un grupo sustituyente de azúcar entre: alquilo C_1 a C_{10} inferior, alquilo inferior sustituido, alquenilo, alquinilo, alcarilo, aralquilo, O-alcarilo u O-aralquilo, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocicloalquilo, heterocicloalcarilo, aminoalquilamino, polialquilamino, sililo sustituido, un grupo de escisión de ARN, un grupo indicador, un intercalante, un grupo para mejorar las propiedades farmacocinéticas de un ácido nucleico dirigido a ácido nucleico, o un grupo para mejorar las propiedades farmacodinámicas de un ácido nucleico dirigido a ácido nucleico y otros sustituyentes que tienen propiedades similares. Una modificación adecuada puede incluir 2'-metoxietoxi (2'-O-CH_2 CH_2OCH_3, también conocido como 2'-O-(2-metoxietilo) o 2'-MOE, es decir, un grupo alcoxialcoxi). Una modificación adecuada adicional puede incluir 2'-dimetilaminooxietoxi, (es decir, un grupo $O(CH_2)_2ON(CH_3)_2$, también conocido como 2'-DMAOE) y 2'-dimetilaminoetoxi-etoxi (también conocido como 2'-O-dimetil-amino-etoxi-etilo o 2'-DMAEOE), es decir, 2'-O-CH_2-O-CH_2-$N(CH_3)_2$.

Otros grupos sustituyentes de azúcar adecuados pueden incluir metoxi (-O-CH_3), aminopropoxi (-O CH_2 CH_2 CH_2NH_2), alilo (-CH_2-CH=CH_2), -O-alilo (-O-CH_2-CH=CH_2) y flúor (F). Los grupos sustituyentes de azúcar 2' pueden estar en la posición arabino (arriba) o ribo (abajo). Una modificación adecuada de 2'-arabino es 2'-F. También se pueden hacer modificaciones similares en otras posiciones del compuesto oligomérico, particularmente, la posición 3' del azúcar del nucleósido 3' terminal o en los nucleótidos enlazados en 2'-5' y la posición 5' del nucleótido 5' terminal. Los compuestos oligoméricos también pueden tener miméticos de azúcar tales como fracciones ciclobutilo en lugar del azúcar pentofuranosilo.

Un ácido nucleico dirigido a ácido nucleico también puede incluir modificaciones o sustituciones de nucleobase (a menudo denominadas simplemente "base"). Como se usa en el presente documento, Las nucleobases "no modificadas" o "naturales" pueden incluir las bases de purina, (por ejemplo, adenina (A) y guanina (G)), y las bases de pirimidina, (por ejemplo, timina (T), citosina (C) y uracilo (U). Las nucleobases modificadas pueden incluir otras nucleobases sintéticas y naturales, tales como 5-metilcitosina (5-me-C), 5-hidroximetil-citosina, xantina, hipoxantina, 2-aminoadenina, derivados de 6-metilo y otros de alquilo de adenina y guanina, derivados de 2-propilo y otros de alquilo de adenina y guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina, 5-propinilo (-C = C-CH_3) uracilo y citosina, y otros derivados alquinílicos de bases de pirimidina, 6-azo uracilo, citosina y timina, 5-uracilo (pseudouracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tioalquilo, 8-hidroxilo y otras adeninas y guaninas 8-sustituidas, 5-halo, particularmente, 5-bromo, 5-trifluorometil y otros uracilos y citosinas 5-sustituidas, 7-metilguanina y 7-metiladenina, 2-F-adenina, 2-aminoadenina, 8-azaguanina y 8-azaadenina, 7-desazaguanina y 7-desazaadenina y 3-desazaguanina y 3-desazaadenina. Las nucleobases modificadas pueden incluir pirimidinas tricíclicas tales como fenoxazina citidina (1H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), fenotiazina citidina (1H-pirimido(5,4-b)(1,4)benzotiazin-2(3H)-ona), La fijación de G tal como una fenoxazina citidina sustituida (por ejemplo, 9-(2-aminoetoxi)-H-pirimido(5,4-(b)(1,4)benzoxazin-2(3H)-ona), carbazol citidina (2H-pirimido(4,5-b)indol-2-ona), piridoindol citidina (H-pirido(3',2':4,5)pirrolo(2,3-d)pirimidin-2-ona).

Las fracciones de bases heterocíclicas pueden incluir aquellas en las que la base de purina o pirimidina se reemplaza por otros heterociclos, por ejemplo, 7-desaza-adenina, 7-desazaguanosina, 2-aminopiridina y 2-piridona. Las nucleobases pueden ser útiles para aumentar la afinidad de unión de un compuesto polinucleotídico. Estas pueden incluir pirimidinas 5-sustituidas, 6-azapirimidinas y purinas N-2, N-6 y O-6-sustituidas, incluyendo 2-aminopropiladenina, 5-propiniluracilo y 5-propinilcitosina. Las sustituciones de 5-metilcitosina pueden aumentar la estabilidad del dúplex de ácido nucleico a 0,6-1,2 ºC y pueden ser sustituciones de bases adecuadas (por ejemplo, cuando se combinan con modificaciones de azúcar 2'-O-metoxietilo).

Una modificación de un ácido nucleico dirigido a ácido nucleico puede comprender unir químicamente al ácido nucleico dirigido a ácido nucleico una o más fracciones o conjugados que puedan potenciar la actividad, distribución celular o absorción celular del ácido nucleico dirigido a ácido nucleico. Estas fracciones o conjugados pueden incluir grupos conjugados unidos covalentemente a grupos funcionales, tales como grupos hidroxilo primarios o secundarios. Los grupos conjugados pueden incluir, pero sin limitación, intercaladores, moléculas indicadoras, poliaminas, poliamidas, polietilenglicoles, poliéteres, grupos que potencian las propiedades farmacodinámicas de los oligómeros y grupos que pueden potenciar las propiedades farmacocinéticas de los oligómeros. Los grupos conjugados pueden incluir, pero sin limitación, colesteroles, lípidos, fosfolípidos, biotina, fenazina, folato, fenantridina, antraquinona, acridina, fluoresceínas, rodaminas, coumarinas y colorantes. Los grupos que potencian las propiedades farmacodinámicas incluyen grupos que mejoran la absorción, potencian la resistencia a la degradación y/o fortalecen la hibridación específica de secuencia con el ácido nucleico diana. Los grupos que pueden potenciar las propiedades farmacocinéticas incluyen grupos que mejoran la absorción, la distribución, el metabolismo o la excreción de un ácido nucleico. Las fracciones de conjugados pueden incluir, pero sin limitación, fracciones lipídicas tales como una fracción de colesterol, ácido cólico un tioéter, (por ejemplo, hexil-S-tritiltiol), un tiocolesterol, una cadena alifática (por ejemplo, restos de dodecandiol o de undecilo), un fosfolípido (por ejemplo, di-hexadecil-rac-glicerol o 1,2-di-0-hexadecil-rac-

glicero-3-H-fosfonato de trietilamonio), una poliamina o una cadena de polietilenglicol o ácido adamantanoacético, una fracción de palmitilo o una fracción de octadecilamina o de hexilamino-carbonil-oxicolesterol.

Una modificación puede incluir un "dominio de transducción de proteínas" o PTD (es decir, un péptido penetrante en células (CPP)). El PTD puede referirse a un polipéptido, polinucleótido, carbohidratos, o compuestos orgánicos o inorgánicos que facilitan atravesar una bicapa lipídica, micela, membrana celular, membrana de orgánulo o membrana vesicular. Un PTD se puede unir a otra molécula, que puede variar desde una molécula polar pequeña hasta una macromolécula grande y/o una nanopartícula, y puede facilitar que la molécula atraviese una membrana, por ejemplo, pasando del espacio extracelular al espacio intracelular, o del citosol al interior de un orgánulo. Un PTD puede estar unido covalentemente al extremo amino de un polipéptido. Un PTD puede estar unido covalentemente al extremo carboxilo de un polipéptido. Un PTD puede estar unido covalentemente a un ácido nucleico. Los ejemplos de PTD pueden incluir, pero sin limitación, un dominio mínimo de transducción de proteínas y péptidos; una secuencia de poliarginina que comprende varias argininas suficientes para dirigir la entrada a una célula (por ejemplo, 3, 4, 5, 6, 7, 8, 9, 10 o 10-50 argininas), un dominio VP22, un dominio de transducción de proteínas de *Drosophila Antennapedia*, un péptido de calcitonina humana truncado, polilisina y transportan, homopolímero de arginina de 3 restos de arginina a 50 restos de arginina. El PTD puede ser un CPP activable (ACPP). Los ACPP pueden comprender un CPP policatiónico (por ejemplo, Arg9 o "R9") conectado a través de un enlazador escindible a un polianión correspondiente (por ejemplo, Glu9 o "E9"), que puede reducir la carga neta a casi cero y, por lo tanto, inhibe la adhesión y la absorción en las células. Tras la escisión del enlazador, el polianión puede ser liberado, desenmascarando localmente la poliarginina y su adhesividad inherente, "activando" así el ACPP para atravesar la membrana.

"Nucleótido" puede referirse, en general, a una combinación de base-azúcar-fosfato. Un nucleótido puede comprender un nucleótido sintético. Un nucleótido puede comprender un análogo de nucleótido sintético. Los nucleótidos pueden ser unidades monoméricas de una secuencia de ácido nucleico (por ejemplo, ácido desoxirribonucleico (ADN) y ácido ribonucleico (ARN)). El término nucleótido puede incluir ribonucleósido trifosfato adenosina trifosfato (ATP), trifosfato de uridina (UTP), trifosfato de citosina (CTP), trifosfato de guanosina (GTP) y trifosfatos de desoxirribonucleósidos tales como dATP, dCTP, dITP, dUTP, dGTP, dTTP o derivados de los mismos. Dichos derivados pueden incluir, por ejemplo, [αS]dATP, 7-desaza-dGTP y 7-desaza-dATP, y derivados de nucleótidos que confieren resistencia a la nucleasa en la molécula de ácido nucleico que los contiene. El término nucleótido, como se usa en el presente documento, puede referirse a didesoxirribonucleósido trifosfatos (ddNTP) y sus derivados. Los ejemplos ilustrativos de didesoxirribonucleósidos trifosfatos pueden incluir, pero sin limitación, ddATP, ddCTP, ddGTP, ddITP y ddTTP. Un nucleótido puede estar sin marcar o marcarse de manera detectable mediante técnicas bien conocidas. El marcaje también se puede realizar con puntos cuánticos. Los marcadores detectables pueden incluir, por ejemplo, isótopos radiactivos, marcadores fluorescentes, marcadores quimioluminiscentes, marcadores bioluminiscentes y marcadores enzimáticos. Los marcadores fluorescentes de nucleótidos pueden incluir, entre otros, fluoresceína, 5-carboxifluoresceína (FAM), 2'7'-dimetoxi-4'5-dicloro-6-carboxifluoresceína (JOE), rodamina, 6-carboxirrodamina (R6G), N,N,N',N'-tetrametil-6-carboxirrodamina (TAMRA), 6-carboxi-X-rodamina (ROX), ácido 4-(4-dimetilaminofenilazo)benzoico (DABCYL), Cascade Blue, Oregon Green, Texas Red, Cianina y ácido 5-(2'-aminoetil)aminonaftalen-1-sulfónico (EDANS). Los ejemplos específicos de nucleótidos marcados con fluorescencia pueden incluir [R6G] dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP y [dROX]ddTTP disponibles en Perkin Elmer, Foster City, Calif FluoroLink DeoxyNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink Fluor X-dCTP, FluoroLink Cy3-dUTP y FluoroLink Cy5-dUTP disponibles en Amersham, Arlington Heights, Ill.; Fluoresceína-15-dATP, Fluoresceína-12-dUTP, Tetrametil-rodamina-6-dUTP, IR770-9-dATP, Fluoresceína-12-ddUTP, Fluoresceína-12-UTP y Fluoresceína-15-2'-dATP disponibles en Boehringer Mannheim, Indianápolis, Ind.; y nucleótidos marcados con cromosomas, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, Cascade Blue-7-UTP, Cascade Blue-7-dUTP, fluoresceína-12-UTP, fluoresceína-12-dUTP, Oregon Green 488-5-dUTP, Rodamina Green-5 -UTP, Rodamina Green-5-dUTP, tetrametilrrodamina-6-UTP, tetrametilrrodamina-6-dUTP, Texas Red-5-UTP, Texas Red-5-dUTP y Texas Red-12-dUTP disponible en Molecular Probes, Eugene, Oreg. Los nucleótidos también se pueden marcar mediante modificación química. Un nucleótido individual modificado químicamente puede ser biotina-dNTP. Algunos ejemplos no limitantes de dNTP biotinilados pueden incluir, biotina-dATP (por ejemplo, bio-N6-ddATP, biotina-14-dATP), biotina-dCTP (por ejemplo, biotina-11-dCTP, biotina-14-dCTP) y biotina-dUTP (por ejemplo, biotina-11-dUTP, biotina-16-dUTP, biotina-20-dUTP).

Como se usa en el presente documento, "Dominio P" puede referirse a una región de un ácido nucleico dirigido a ácido nucleico. El dominio P puede interactuar con un motivo adyacente al protoespaciador (PAM), polipéptido dirigido a un sitio y/o ácido nucleico dirigido a ácido nucleico. Un dominio P puede interactuar directa o indirectamente con un motivo adyacente al protoespaciador (PAM), polipéptido dirigido a un sitio y/o ácido nucleico dirigido a ácido nucleico ácido nucleico. Como se usan en el presente documento, las expresiones "región interactiva de PAM", "región adyacente anti-repetición" y "dominio P" se pueden usar indistintamente.

Como se usa en el presente documento, "purificado/a" puede referirse a una molécula (por ejemplo, polipéptido dirigido a un sitio, ácido nucleico dirigido a ácido nucleico) que comprende al menos el 50, 60, 70, 80, 90, 95, 96, 97, 98, 99 o 100 % de la composición. Por ejemplo, si una muestra que comprende el 10 % de un polipéptido dirigido a un sitio, después de una etapa de purificación comprende el 60 % del polipéptido dirigido a un sitio, entonces se puede decir

que la muestra está purificada. Una muestra purificada puede referirse a una muestra enriquecida o una muestra que ha sido sometida a métodos para eliminar partículas distintas de la partícula de interés.

Como se usa en el presente documento, "agente de reactivación" puede referirse, en general, a cualquier agente que pueda convertir un polipéptido inactivo enzimáticamente en un polipéptido activo enzimáticamente. El imidazol puede ser un agente de reactivación. Un análogo de ligando puede ser un agente de reactivación.

Como se usa en el presente documento, "recombinante" puede referirse a una secuencia que procede de una fuente foránea al hospedador en particular (por ejemplo, célula) o, si procede de la misma fuente, está modificada con respecto a su forma original. Un ácido nucleico recombinante de una célula puede incluir un ácido nucleico que sea endógeno a la célula en particular, pero que se haya modificado a través de, por ejemplo, el uso de mutagénesis dirigida. El término puede incluir múltiples copias no naturales de una secuencia de ADN que se produce de forma natural. Así pues, el término puede referirse a un ácido nucleico que sea foráneo o heterólogo a la célula, u homólogo a la célula, pero en una posición o forma dentro de la célula en la que normalmente no se encuentra el ácido nucleico. De igual manera, cuando se usa en el contexto de una secuencia polipeptídica o de aminoácidos, una secuencia polipeptídica o de aminoácidos exógena puede ser una secuencia polipeptídica o de aminoácidos que proceda de una fuente foránea para la célula en particular o, si procede de la misma fuente, está modificada con respecto a su forma original.

Como se usa en el presente documento, "polipéptidos dirigidos a un sitio" puede referirse, en general, a nucleasas, nucleasas dirigidas a un sitio, endorribonucleasas, endorribonucleasas condicionalmente inactivas enzimáticamente, proteínas argonautas y proteínas de unión a ácidos nucleico. Un polipéptido o una proteína dirigido a un sitio puede incluir nucleasas tales como endonucleasas de asentamiento tales como PI-TliII, H-DreI, I-DmoI y I-CreI, I-SceI, nucleasas de la familia LAGLIDADG, meganucleasas, nucleasas de la familia GIY-YIG, nucleasas de la familia de cajas His-Cys, nucleasas de tipo Vsr, endorribonucleasas, exorribonucleasas, endonucleasas y exonucleasas. Un polipéptido dirigido a un sitio puede referirse a un miembro de los genes Cas de los sistemas de CRISPR/Cas de Tipo I, Tipo II, Tipo III y/o Tipo U. Un polipéptido dirigido a un sitio puede referirse a un miembro de la superfamilia de proteínas misteriosas asociadas a repeticiones (RAMP) (por ejemplo, subfamilias Cas5, Cas6). Un polipéptido dirigido a un sitio puede referirse a una proteína argonauta.

Un polipéptido dirigido a un sitio puede ser un tipo de proteína. Un polipéptido dirigido a un sitio puede referirse a una nucleasa. Un polipéptido dirigido a un sitio puede referirse a una endorribonucleasa. Un polipéptido dirigido a un sitio puede referirse a cualquier secuencia polipeptídica modificada (por ejemplo, acortada, mutada, alargada) u homólogo del polipéptido dirigido a un sitio. Un polipéptido dirigido a un sitio puede ser codón optimizado. Un polipéptido dirigido a un sitio puede ser un homólogo optimizado por codones de un polipéptido dirigido a un sitio. Un polipéptido dirigido a un sitio puede ser enzimáticamente inactivo, parcialmente activo, constitutivamente activo, completamente activo, inducible activo y/o más activo, (por ejemplo, más que el homólogo de tipo silvestre de la proteína o del polipéptido). Un polipéptido dirigido a un sitio puede ser Cas9. Un polipéptido dirigido a un sitio puede ser Csy4. Un polipéptido dirigido a un sitio puede ser Cas5 o un miembro de la familia Cas5. Un polipéptido dirigido a un sitio puede ser Cas6 o un miembro de la familia Cas6.

En algunos casos, el polipéptido dirigido a un sitio (por ejemplo, polipéptido dirigido a un sitio variante, mutado, inactivo enzimáticamente y/o condicionalmente inactivo enzimáticamente) puede dirigirse a ácido nucleico. El polipéptido dirigido a un sitio (por ejemplo, endorribonucleasa variante, mutada, enzimáticamente inactiva y/o condicionalmente enzimáticamente inactiva) puede dirigirse a ARN. Las endorribonucleasas que pueden dirigirse a ARN pueden incluir miembros de otras subfamilias de CRISPR tales como Cas6 y Cas5.

Como se usa en el presente documento, el término "específico" puede referirse a la interacción de dos moléculas a través de la que una de las moléculas, a través de, por ejemplo, medios químicos o físicos, se une específicamente a la segunda molécula. Las interacciones de unión específicas ilustrativas pueden referirse a la unión de antígeno-anticuerpo, unión de avidina-biotina, carbohidratos y lectinas, secuencias de ácido nucleico complementarias (por ejemplo, hibridación), secuencias peptídicas complementarias, incluyendo las formadas mediante métodos recombinantes, moléculas efectoras y receptoras, cofactores enzimáticos y enzimas, inhibidores enzimáticos y enzimas, y similares. "No específica" puede referirse a una interacción entre dos moléculas que no es específica.

Como se usa en el presente documento, "soporte sólido" puede referirse, en general, a cualquier material insoluble o parcialmente soluble. Un soporte sólido puede referirse a una tira reactiva, una placa de múltiples pocillos y similares. El soporte sólido puede comprender varias sustancias (por ejemplo, vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poliacrilamidas, agarosas y magnetita), y se pueden proporcionar en varias formas, incluyendo perlas de agarosa, perlas de poliestireno, perlas de látex, perlas magnéticas, partículas metálicas coloidales, chips y superficies de vidrio y/o silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de bandejas de reacción (por ejemplo, placas de múltiples pocillos), tubos de plástico, etc. Un soporte sólido puede ser sólido, semisólido, una perla o una superficie. El soporte puede ser móvil en una solución o puede ser inmóvil. Se puede usar un soporte sólido para capturar un polipéptido. Un soporte sólido puede comprender un agente de captura.

Como se usa en el presente documento, "ácido nucleico diana" puede referirse, en general, a un ácido nucleico que se vaya a usar en los métodos de la divulgación. Un ácido nucleico diana puede referirse a una secuencia cromosómica o una secuencia extracromosómica, (por ejemplo, una secuencia episomal, una secuencia de minicírculo, una secuencia mitocondrial, una secuencia de cloroplasto, etc.). Un ácido nucleico diana puede ser ADN. Un ácido nucleico diana puede ser ARN. Un ácido nucleico diana puede usarse en el presente documento indistintamente con "polinucleótido", "secuencia de nucleótidos" y/o "polinucleótido diana". Un ácido nucleico diana puede ser una secuencia de ácido nucleico que puede no estar relacionada con ninguna otra secuencia de una muestra de ácido nucleico mediante una sustitución de un solo nucleótido. Un ácido nucleico diana puede ser una secuencia de ácido nucleico que puede no estar relacionada con ninguna otra secuencia de una muestra de ácido nucleico mediante sustituciones de 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos. En algunas realizaciones, la sustitución no puede ocurrir dentro de los 5, 10, 15, 20, 25, 30 o 35 nucleótidos del extremo 5' de un ácido nucleico diana. En algunas realizaciones, la sustitución no puede ocurrir dentro de 5, 10, 15, 20, 25, 30, 35 nucleótidos del extremo 3' de un ácido nucleico diana.

Como se usa en el presente documento, "ARNcrtra" puede referirse, en general, a un ácido nucleico con al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con una secuencia de ARNcrtra ilustrativa de tipo silvestre (por ejemplo, un ARNcrtra de *S. pyogenes* (SEQ ID 433), SEQ ID 431-562). ARNcrtra puede referirse a un ácido nucleico con, como máximo, aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % de identidad de secuencia y/o similitud de secuencia con una secuencia de ARNcrtra ilustrativa de tipo silvestre (por ejemplo, un ARNcrtra de *S. pyogenes*). ARNcrtra puede referirse a una forma modificada de un ARNcrtra que puede comprender un cambio de nucleótidos tal como una eliminación, inserción o sustitución, variante, mutación o quimera. Un ARNcrtra puede referirse a un ácido nucleico que puede ser al menos aproximadamente un 60 % idéntico a una secuencia de ARNcrtra ilustrativa de tipo silvestre (por ejemplo, un ARNcrtra de *S. pyogenes*) en un tramo de al menos 6 nucleótidos contiguos. Por ejemplo, una secuencia de ARNcrtra puede ser al menos aproximadamente un 60 % idéntica, al menos aproximadamente un 65 % idéntica, al menos aproximadamente un 70 % idéntica, al menos aproximadamente un 75 % idéntica, al menos aproximadamente un 80 % idéntica, al menos aproximadamente un 85 % idéntica, al menos aproximadamente un 90 % idéntico, al menos aproximadamente un 95 % idéntica, al menos aproximadamente un 98 % idéntica, al menos aproximadamente un 99 % idéntica o un 100 % idéntica, a una secuencia de ARNcrtra ilustrativa de tipo silvestre (por ejemplo, un ARNcrtra de *S. pyogenes*) en un tramo de al menos 6 nucleótidos contiguos. Un ARNcrtra puede referirse a un ARNcrtra medio. Un ARNcrtra puede referirse a una secuencia mínima de ARNcrtra.

Sistemas de CRISPR

Un CRISPR (grupos de repeticiones palindrómicas cortas en intervalos regulares) puede ser un locus genómico que se encuentra en los genomas de muchos procariotas (por ejemplo, bacterias y arqueas). Los locus de CRISPR pueden proporcionar resistencia a los invasores foráneos (por ejemplo, virus, fago) en procariotas. De esta manera, se puede pensar que el sistema de CRISPR funciona como un tipo de sistema inmunitario para ayudar a defender a los organismos procariotas contra los invasores foráneos. Puede haber tres etapas de la función del locus de CRISPR: integración de nuevas secuencias en el locus, biogénesis del ARN de CRISPR (ARNcr) y silenciamiento del ácido nucleico invasor foráneo. Puede haber cuatro tipos de sistemas de CRISPR (por ejemplo, Tipo I, Tipo II, Tipo III, Tipo U).

Un locus de CRISPR puede incluir una serie de secuencias repetitivas cortas denominadas "repeticiones". Las repeticiones pueden formar estructuras de horquilla y/o las repeticiones pueden ser secuencias monocatenarias no estructuradas. Las repeticiones pueden ocurrir en grupos. Las secuencias de repeticiones, con frecuencia, pueden divergir entre especies. Las repeticiones pueden estar intercaladas regularmente con secuencias intermedias únicas denominadas "espaciadores", produciendo una arquitectura de locus de repetición-espaciador-repetición. Los espaciadores pueden ser idénticos o tener una alta homología con secuencias invasoras foráneas conocidas. Una unidad de espaciador-repetición puede codificar un ARNcrispr (ARNcr). Un ARNcr puede referirse a la forma madura de la unidad de espaciador-repetición. Un ARNcr puede comprender una secuencia "semilla" que puede participar en la dirección a un ácido nucleico diana (por ejemplo, posiblemente como un mecanismo de vigilancia contra ácido nucleico foráneo). Se puede ubicar una secuencia semilla en el extremo 5' o 3' del ARNcr.

Un locus de CRISPR puede comprender secuencias polinucleotídicas que codifican genes asociados a Crispr (Cas). Los genes Cas pueden participar en las etapas de biogénesis y/o interferencia de la función de ARNcr. Los genes Cas pueden mostrar una divergencia de secuencia extrema (por ejemplo, secuencia primaria) entre especies y homólogos. Por ejemplo, los homólogos Cas1 pueden comprender menos del 10 % de identidad con la secuencia primaria entre homólogos. Algunos genes Cas pueden comprender estructuras secundarias y/o terciarias homólogas. Por ejemplo, a pesar de la divergencia de secuencia extrema, muchos miembros de la familia Cas6 de proteínas CRISPR comprenden un pliegue de tipo ferredoxina N-terminal. Los genes Cas se pueden nombrar de acuerdo con el organismo del que se derivan. Por ejemplo, los genes Cas de *Staphylococcus epidermidis* pueden denominarse de tipo Csm, los genes Cas de *Streptococcus thermophilus* pueden denominarse de tipo Csn y los genes Cas de *Pyrococcus furiosus* pueden denominarse de tipo Cmr.

Integración

La etapa de integración del sistema de CRISPR puede referirse a la capacidad del locus de CRISPR para integrar nuevos espaciadores en la matriz de ARNcr al ser infectado por un invasor foráneo. La adquisición de espaciadores invasores foráneos puede ayudar a conferir inmunidad a los ataques posteriores del mismo invasor foráneo. La integración puede ocurrir en el extremo líder del locus de CRISPR. Las proteínas Cas (por ejemplo, Cas1 y Cas2) pueden participar en la integración de nuevas secuencias de espaciador. La integración puede proceder de manera similar para algunos tipos de sistemas de CRISPR (por ejemplo, Tipo I-III).

Biogénesis

Los ARNcr maduros pueden procesarse a partir de una transcripción de locus de CRISPR policistrónica más larga (es decir, una matriz de pre-ARNcr). Una matriz de pre-ARNcr puede comprender una pluralidad de ARNcr. Las repeticiones de la matriz de pre-ARNcr pueden ser reconocidas por los genes Cas. Los genes Cas pueden unirse a las repeticiones y escindir las repeticiones. Esta acción puede liberar la pluralidad de ARNcr. Los ARNcr pueden someterse a eventos adicionales para producir la forma madura de ARNcr, tales como el recorte (por ejemplo, con un exonucleasa). Un ARNcr puede comprender todas, alguna o ninguna de la secuencia de repetición de CRISPR.

Interferencia

La interferencia puede referirse a la etapa del sistema de CRISPR que es funcionalmente responsable de combatir la infección por un invasor foráneo. La interferencia de CRISPR puede seguir un mecanismo similar a la interferencia del ARN (ARNi (por ejemplo, en donde un ARN diana está dirigido (por ejemplo, hibridado) por un ARN interferente corto (ARNip)), que puede producir la degradación y/o desestabilización del ARN diana. Los sistemas de CRISPR pueden realizar LA interferencia de un ácido nucleico diana mediante el acoplamiento de ARNcr y genes Cas, formando así ribonucleoproteínas de CRISPR (RNPcr). El ARNcr de la RNPcr puede guiar al RNPcr al ácido nucleico invasor foráneo, (por ejemplo, reconociendo el ácido nucleico invasor foráneo a través de hibridación). Las unidades de ARNcr de ácido nucleico invasor foráneo diana hibridado pueden ser sometidas a escisión por proteínas Cas. La interferencia del ácido nucleico diana puede requerir un motivo adyacente al espaciador (PAM) en un ácido nucleico diana.

Tipos de sistemas de CRISPR

Puede haber cuatro tipos de sistemas de CRISPR: Tipo I, Tipo II, Tipo III y Tipo U. Se puede encontrar más de un tipo de sistema de CRISPR en un organismo. Los sistemas de CRISPR pueden ser complementarios entre sí y/o pueden prestar unidades funcionales en *trans* para facilitar el procesamiento del locus de CRISPR.

Sistemas de CRISPR de tipo I

La biogénesis del ARNcr en los sistemas de CRISPR de Tipo I puede comprender la escisión por endorribonucleasa de repeticiones de la matriz de pre-ARNcr, que puede producir una pluralidad de ARNcr. Los ARNcr de los sistemas de Tipo I no pueden someterse a recorte de ARNcr. Un ARNcr puede procesarse a partir de una matriz de pre-ARNcr mediante un complejo multiproteína denominado Cascade (que se origina a partir del complejo asociado a CRISPR para la defensa antivírica). Cascade puede comprender subunidades de proteínas (por ejemplo, CasA-CasE). Algunas de las subunidades pueden ser miembros de la superfamilia de proteínas misteriosas asociadas a repeticiones (RAMP) (por ejemplo, familias Cas5 y Cas6). El complejo de Cascade-ARNcr (es decir, complejo de interferencia) puede reconocer el ácido nucleico diana a través de la hibridación del ARNcr con el ácido nucleico diana. El complejo de interferencia de Cascade puede reclutar la helicasa/nucleasa Cas3 que puede actuar en *trans* para facilitar la escisión del ácido nucleico diana. La nucleasa Cas3 puede escindir el ácido nucleico diana (por ejemplo, con su dominio nucleasa HD). El ácido nucleico diana de un sistema de CRISPR de Tipo I puede comprender un PAM. El ácido nucleico diana de un sistema de CRISPR DE Tipo I puede ser ADN.

Los sistemas de Tipo I pueden subdividirse aún más por su especie de origen. Los sistemas de Tipo I pueden comprender: las subfamilias de los Tipos IA (*Aeropyrum pernix* o CASS5); IB (*Thermotoga neapolitana- Haloarcula marismortui* o CASS7); IC (*Desulfovibrio vulgaris* o CASS1); ID; IE *(Escherichia coli* o CASS2); e IF (*Yersinia pestis* o CASS3).

Sistemas de CRISPR de Tipo II

La biogénesis de ARNcr en un sistema de CRISPR de Tipo II puede comprender un ARN de CRISPR de activación en *trans* (ARNcrtra). Un ARNcrtra puede ser modificado por RNasaIII endógena. El ARNcrtra del complejo puede hibridarse con una repetición de ARNcr en la matriz de pre-ARNcr. RnasaIII endógena puede ser reclutada para escindir el pre-ARNcr. Los ARNcr escindidos pueden someterse a un recorte de exorribonucleasa para producir la forma madura de ARNcr (por ejemplo, recorte de 5'). El ARNcrtra puede permanecer hibridado con el ARNcr. El ARNcrtra y el ARNcr pueden asociarse con un polipéptido dirigido a un sitio (por ejemplo, Cas9). El ARNcr del complejo de ARNcr-ARNcrtra-Cas9 puede guiar al complejo a un ácido nucleico diana con el que el ARNcr puede hibridarse. La hibridación del ARNcr con el ácido nucleico diana puede activar a Cas9 para la escisión del ácido nucleico diana. El ácido nucleico diana de un sistema de CRISPR de Tipo II puede comprender un PAM. En algunas realizaciones, un PAM es esencial para facilitar la unión de un polipéptido dirigido a un sitio (por ejemplo, Cas9) a un ácido nucleico

diana. Los sistemas de Tipo II se pueden subdividir en II-A (Nmeni o CASS4) y II-B (Nmeni o CASS4).

Sistemas de CRISPR de Tipo III

La biogénesis de ARNcr de los sistemas de CRISPR de Tipo III puede comprender una etapa de escisión por endorribonucleasa de repeticiones de la matriz de pre-ARNcr, que puede producir una pluralidad de ARNcr. Las repeticiones del sistema de CRISPR de Tipo III pueden ser regiones monocatenarias no estructuradas. Las repeticiones pueden ser reconocidas y escindidas por un miembro de la superfamilia de endorribonucleasas RAMP (por ejemplo, Cas6). Los ARNcr de los sistemas de Tipo III (por ejemplo, Tipo III-B) pueden someterse a recorte de ARNcr (por ejemplo, recorte de 3'). Los sistemas de Tipo III pueden comprender una proteína similar a la polimerasa (por ejemplo, Cas10). Cas10 puede comprender un dominio homólogo a un dominio de palma.

Los sistemas de Tipo III pueden procesar pre-ARNcr con un complejo que comprende una pluralidad de proteínas que forman parte de la superfamilia RAMP y una o más proteínas CRISPR de tipo polimerasa. Los sistemas de Tipo III se pueden dividir en III-A y III-B. Un complejo de interferencia del sistema DE Tipo III-A (es decir, complejo Csm) puede dirigirse a ácido nucleico plasmídico. La escisión del ácido nucleico plasmídico puede ocurrir con el dominio nucleasa HD de una proteína similar a la polimerasa en el complejo. Un complejo de interferencia del sistema de Tipo III-B (es decir, complejo Cmr) puede dirigirse a ARN.

Sistemas de CRISPR de Tipo U

Los sistemas de CRISPR de Tipo U pueden no comprender los genes distintivos de ninguno de los sistemas de CRISPR de Tipo I-III (por ejemplo, Cas3, Cas9, Cas6, Cas1, Cas2). Los ejemplos de genes Cas de CRISPR de Tipo U pueden incluir, pero sin limitación, Csf1, Csf2, Csf3, Csf4. Los genes Cas de Tipo U pueden ser homólogos muy distantes de los genes Cas de Tipo I-III. Por ejemplo, Csf3 puede ser muy divergente, pero funcionalmente similar a los miembros de la familia Cas5. Un sistema de Tipo U puede funcionar complementariamente en *trans* con un sistema de Tipo I-III. En algunos casos, los sistemas de tipo U pueden no estar asociados con el procesamiento de matrices de CRISPR. Los sistemas de Tipo U pueden representar un sistema alternativo de defensa contra invasores foráneos.

Superfamilia de RAMP

Las proteínas misteriosas asociadas a repeticiones (proteínas RAMP) se pueden caracterizar por un pliegue proteico que comprende un motivo βαββαβ [beta-alfa-beta-beta-alfa-beta] de cadenas β (β) y hélices α (α). Una proteína RAMP puede comprender un motivo de reconocimiento de ARN (RRM) (que puede comprender una ferredoxina o un pliegue de tipo ferredoxina). Las proteínas RAMP pueden comprender un RRM N-terminal. El dominio C-terminal de las proteínas RAMP puede variar, pero también puede comprender un RRM. Los miembros de la familia RAMP pueden reconocer el ácido nucleico estructurado y/o no estructurado. Los miembros de la familia RAMP pueden reconocer el ácido nucleico monocatenario y/o bicatenario. Las proteínas RAMP pueden participar en la biogénesis y/o la etapa de interferencia de los sistemas de CRISPR de Tipo I y Tipo III. Los miembros de la superfamilia RAMP pueden comprender miembros de familias Cas7, Cas6 y Cas5. Los miembros de la superfamilia RAMP pueden ser endorribonucleasas.

Los dominios RRM de la superfamilia RAMP pueden ser extremadamente divergentes. Los dominios RRM pueden comprender al menos aproximadamente un 5 %, al menos aproximadamente un 10 %, al menos aproximadamente un 15 %, al menos aproximadamente un 20 %, al menos aproximadamente un 25 %, al menos aproximadamente un 30 %, al menos aproximadamente un 35 %, al menos aproximadamente un 40 %, al menos aproximadamente un 45 %, al menos aproximadamente un 50 %, al menos aproximadamente un 55 %, al menos aproximadamente un 60 %, al menos aproximadamente un 65 %, al menos aproximadamente un 70 %, al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 % o 100 % de homología de secuencia o estructural con un dominio RRM ilustrativo de tipo silvestre (por ejemplo, un dominio RRM de Cas7). Los dominios RRM pueden comprender, como máximo, aproximadamente un 5 %, como máximo, aproximadamente un 10 %, como máximo, aproximadamente un 15 %, como máximo, aproximadamente un 20 %, como máximo, aproximadamente un 25 %, como máximo, aproximadamente un 30 %, como máximo, aproximadamente un 35 %, como máximo, aproximadamente un 40 %, como máximo, aproximadamente un 45 %, como máximo, aproximadamente un 50 %, como máximo, aproximadamente un 55 %, como máximo, aproximadamente un 60 %, como máximo, aproximadamente un 65 %, como máximo, aproximadamente un 70 %, como máximo, aproximadamente un 75 %, como máximo, aproximadamente un 80 %, como máximo, aproximadamente un 85 %, como máximo, aproximadamente un 90 %, como máximo aproximadamente un 95 % o 100 % de homología de secuencia o estructural con un dominio RRM ilustrativo de tipo silvestre (por ejemplo, un dominio RRM de Cas7).

Familia Cas7

Los miembros de la familia Cas7 pueden ser una subclase de proteínas de la familia RAMP. Las proteínas de la familia Cas7 se pueden clasificar en sistemas de CRISPR de Tipo I. Los miembros de la familia Cas7 pueden no comprender un bucle rico en glicina que sea familiar para algunos miembros de la familia RAMP. Los miembros de la familia Cas7

pueden comprender un dominio RRM. Los miembros de la familia Cas7 pueden incluir, pero sin limitación, Cas7 (COG1857), Cas7 (COG3649), Cas7 (CT1975), Csy3, Csm3, Cmr6, Csm5, Cmr4, Cmr1, Csf2 y Csc2.

Familia Cas6

La familia Cas6 puede ser una subfamilia de RAMP. Los miembros de la familia Cas6 pueden comprender dos dominios similares a motivos de reconocimiento de ARN (RRM). Un miembro de la familia Cas6 (por ejemplo, Cas6f) puede comprender un dominio RRM N-terminal y un dominio C-terminal distinto que puede mostrar una similitud de secuencia u homología estructural débil con un dominio RRM. Los miembros de la familia Cas6 pueden comprender una histidina catalítica que puede participar en la actividad endorribonucleasa. Se puede encontrar un motivo comparable en las familias de RAMP Cas5 y Cas7. Los miembros de la familia Cas6 pueden incluir, pero sin limitación, Cas6, Cas6e, Cas6f (por ejemplo, Csy4).

Familia Cas5

La familia Cas5 puede ser una subfamilia de RAMP. La familia Cas5 se puede dividir en dos subgrupos: un subgrupo que puede comprender dos dominios RRM y un subgrupo que puede comprender un dominio RRM. Los miembros de la familia Cas5 pueden incluir, pero sin limitación, Csm4, Csx10, Cmr3, Cas5, Cas5(BH0337), Csy2, Csc1, Csf3.

Genes Cas

Los genes Cas de CRISPR ilustrativos pueden incluir Cas1, Cas2, Cas3' (cebado de Cas3), Cas3'' (cebado doble de Cas3), Cas4, Cas5, Cas6, Cas6e (anteriormente denominado CasE, Cse3), Cas6f (es decir, Csy4), Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4. La Tabla 1 proporciona una categorización ilustrativa de genes Cas de CRISPR según el tipo de sistema de CRISPR.

El sistema de nombramiento de los genes Cas de CRISPR ha experimentado una amplia reestructuración desde que se descubrieron los genes Cas. Para los fines de la presente solicitud, los nombres de los genes Cas usados en el presente documento se basan en el sistema de nombramiento descrito en Makarova *et al.* "Evolution and classification of the CRISPR-Cas systems". *Nature Reviews Microbiology*. Junio de 2011; 9(6): 467-477. Doi:10.1038/nrmicro2577.

Tabla 1: Clasificación ilustrativa de genes Cas de CRISPR según el tipo de CRISPR

Tipo o subtipo del sistema	**Nombre del gen**
Tipo I	***cas1, cas2, cas3'***
Tipo II	***cas1, cas2, cas9***
Tipo III	***cas1, cas2, cas10***
Subtipo I-A	***cas3'', cas4, cas5, cas6, cas7, cas8a1, cas8a2, csa5***
Subtipo I-B	***cas3'', cas4, cas5, cas6, cas7, cas8b***
Subtipo I-C	***cas4, cas5, cas7, cas8c***
Subtipo I-D	***cas4, cas6, cas10d, csc1, csc2***
Subtipo I-E	***cas5, cas6e, cas7, cse1, cse2***
Subtipo I-F	***cas6f, csy1, csy2, csy3***
Subtipo II-A	***csn2***
Subtipo II-B	***cas4***
Subtipo III-A	***cas6, csm2, csm3, csm4,*** csm5, csm6
Subtipo III-B	***cas6, cmr1, cmr3, cmr4, cmr5, cmr6***
Subtipo I-U	***csb1, csb2, csb3, csx17, csx14, csx10***
Subtipo III-U	***csx16, csaX, csx3, csx1***
Desconocido	***csx15***
Tipo U	***csf1, csf2, csf3, csf4***

Polipéptidos dirigidos a un sitio

Un polipéptido dirigido a un sitio puede ser un polipéptido que puede unirse a un ácido nucleico diana. Un polipéptido dirigido a un sitio puede ser una nucleasa.

Un polipéptido dirigido a un sitio puede comprender un dominio de unión a ácido nucleico. El dominio de unión a ácido nucleico puede comprender una región que está en contacto con un ácido nucleico. Un dominio de unión a ácido

nucleico puede comprender un ácido nucleico. Un dominio de unión a ácido nucleico puede comprender un material proteico. Un dominio de unión a ácido nucleico puede comprender ácido nucleico y un material proteico. Un dominio de unión a ácido nucleico puede comprender ARN. Puede haber un solo dominio de unión a ácido nucleico. Los ejemplos de dominios de unión a ácido nucleico pueden incluir, pero sin limitación, un dominio hélice-vuelta-hélice, un dominio de dedo de cinc, un dominio de cremallera de leucina (bZIP), un dominio de hélice alada, un dominio de hélice alada-vuelta-hélice, un dominio hélice-lazo-hélice, un dominio de caja HMG, un dominio Wor3, un dominio de inmunoglobulina, un dominio B3, un dominio TALE, un dominio de motivo de reconocimiento de ARN, un dominio de motivo de unión a ARN bicatenario, un dominio de unión a ácido nucleico bicatenario, un dominio de unión a ácido nucleico monocatenario, un dominio KH, un dominio PUF, un dominio de caja RGG, un dominio de caja DEAD/DEAH, un dominio PAZ, un dominio Piwi y un dominio de choque frío.

Un dominio de unión a ácido nucleico puede ser un dominio de una proteína argonauta. Una proteína argonauta puede ser una argonauta eucariota o una argonauta procariota. Una proteína argonauta puede unirse a ARN, ADN, o tanto a ARN como a ADN. Una proteína argonauta puede escindir ARN o ADN, o tanto ARN como ADN. En algunos casos, una proteína argonauta se une a un ADN y escinde un ADN diana.

En algunos casos, dos o más dominios de unión a ácido nucleico pueden estar unidos entre sí. La unión de una pluralidad de dominios de unión a ácido nucleico entre sí puede proporcionar una mayor especificidad dirigida a un polinucleótido. Dos o más dominios de unión a ácido nucleico pueden unirse mediante uno o más enlazadores. El enlazador puede ser un enlazador flexible. Los enlazadores pueden comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40 o más aminoácidos de longitud. Los enlazadores pueden comprender al menos el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de contenido de glicina. Los enlazadores pueden comprender, como máximo, el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de contenido de glicina. Los enlazadores pueden comprender al menos el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de contenido de serina. Los enlazadores pueden comprender, como máximo, el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de contenido de serina.

Los dominios de unión a ácido nucleico pueden unirse a secuencias de ácido nucleico. Los dominios de unión a ácido nucleico pueden unirse a ácidos nucleicos mediante hibridación. Los dominios de unión a ácido nucleico pueden ser diseñados (por ejemplo, diseñados para hibridarse con una secuencia en un genoma). Un dominio de unión a ácido nucleico puede diseñarse mediante técnicas de clonación molecular (por ejemplo, evolución dirigida, mutación específica y mutagénesis racional).

Un polipéptido dirigido a un sitio puede comprender un dominio de escisión de ácido nucleico. El dominio de escisión de ácido nucleico puede ser un dominio de escisión de ácido nucleico de cualquier proteína de escisión de ácido nucleico. El dominio de escisión de ácido nucleico puede proceder de una nucleasa. Los dominios de escisión de ácido nucleico adecuados incluyen el dominio de escisión de ácido nucleico de endonucleasas (por ejemplo, endonucleasa AP, enonucleasa RecBCD, endonucleasa T7, endonucleasa T4 IV, endonucleasa Bal 31, Endonucleasa I (endo I), Nucleasa micrococica, Endonucleasa II (endo VI, exo III)), exonucleasas, nucleasas de restricción, endorribonucleasas, exorribonucleasas, RNasas (por ejemplo, ARNasa I, II o III). En algunos casos, el dominio de escisión de ácido nucleico puede proceder de la endonucleasa FokI. Un polipéptido dirigido a un sitio puede comprender una pluralidad de dominios de escisión de ácido nucleico. Los dominios de escisión de ácido nucleico pueden estar unidos entre sí. Dos o más dominios de escisión de ácido nucleico se pueden unir mediante un enlazador. En algunas realizaciones, el enlazador puede ser un enlazador flexible. Los enlazadores pueden comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40 o más aminoácidos de longitud. En algunas realizaciones, un polipéptido dirigido a un sitio puede comprender la pluralidad de dominios de escisión de ácido nucleico.

Un polipéptido dirigido a un sitio (por ejemplo, Cas9, argonauta) puede comprender dos o más dominios nucleasa. Cas9 puede comprender un dominio nucleasa de tipo HNH o HNH y/o un dominio nucleasa de tipo RuvC o RuvC. Los dominios de tipo HNH o HNH pueden comprender un pliegue de tipo McrA. Los dominios de tipo HNH o HNH pueden comprender dos cadenas β paralelas y una hélice α. Los dominios de tipo HNH o HNH pueden comprender un sitio de unión a metal (por ejemplo, sitio de unión a cationes divalentes). Los dominios de tipo HNH o HNH pueden escindir una cadena de un ácido nucleico diana (por ejemplo, cadena complementaria de la cadena dirigida a ARNcr). Las proteínas que comprenden un dominio de tipo HNH o HNH pueden incluir endonucleasas, clicinas, endonucleasas de restricción, transposasas y factores de empaquetamiento de ADN.

Los dominios RuvC o de tipo RuvC pueden comprender un pliegue de RNasaH o de tipo RNasaH. Los dominios RuvC/RNasaH pueden participar en un conjunto diverso de funciones basadas en ácidos nucleicos, incluyendo la actuación sobre ARN y ADN. El dominio RNasaH puede comprender 5 cadenas β rodeadas por una pluralidad de hélices α. Los dominios RuvC/RNasaH o de tipo RuvC/RNasaH pueden comprender un sitio de unión a metal (por ejemplo, sitio de unión a cationes divalentes). Los dominios RuvC/RNasaH o de tipo RuvC/RNasaH pueden escindir una cadena de un ácido nucleico diana (por ejemplo, cadena no complementaria de la cadena dirigida a ARNcr). Las proteínas que comprenden un dominio RuvC, de tipo RuvC o de tipo RNasaH pueden incluir RNasaH, RuvC,

transposasas de ADN, integrasas retrovíricas y proteínas argonautas).

El polipéptido dirigido a un sitio puede ser una endorribonucleasa. El polipéptido dirigido a un sitio puede ser un polipéptido dirigido a un sitio inactivo enzimáticamente. El polipéptido dirigido a un sitio puede ser un polipéptido dirigido a un sitio condicionalmente inactivo enzimáticamente. Los polipéptidos dirigidos a un sitio pueden introducir roturas bicatenarias o roturas monocatenarias en ácido nucleico, (por ejemplo, ADN genómico). La rotura bicatenaria puede estimular las vías de reparación del ADN endógenas de una célula (por ejemplo, recombinación homóloga y unión de extremos no homólogos (NHEJ) o unión de extremos no homólogos alternativa (A-NHEJ)). La NHEJ puede reparar el ácido nucleico escindido sin la necesidad de una plantilla homóloga. Esto puede producir eliminaciones del ácido nucleico diana. La recombinación homóloga (HR) puede ocurrir con una plantilla homóloga. La plantilla homóloga puede comprender secuencias que son homólogas a las secuencias que flanquean el sitio de escisión del ácido nucleico diana. Una vez escindido un ácido nucleico diana por un polipéptido dirigido a un sitio, el sitio de escisión puede destruirse (por ejemplo, el sitio puede no ser accesible para otra serie de escisión con el ácido nucleico dirigido a ácido nucleico y el polipéptido dirigido a un sitio originales).

En algunos casos, la recombinación homóloga puede introducir una secuencia polinucleotídica exógena en el sitio de escisión del ácido nucleico diana. Una secuencia polinucleotídica exógena puede denominarse polinucleótido donante. En algunos casos de los métodos de divulgación, se puede insertar el polinucleótido donante, una parte de un polinucleótido donante, una copia del polinucleótido donante o una parte de la copia del polinucleótido donante en el sitio de escisión del ácido nucleico diana. Un polinucleótido donante puede ser una secuencia polinucleotídica exógena. Un polinucleótido donante puede ser una secuencia no natural en el sitio de escisión del ácido nucleico diana. Un vector puede comprender un polinucleótido donante. Las modificaciones del ADN diana debidas a la NHEJ y/o HR pueden conducir a, por ejemplo, mutaciones, eliminaciones, alteraciones, integraciones, corrección genética, reemplazo de genes, marcaje de genes, inserción de transgenes, eliminación de nucleótidos, disrupción genética y/o mutación genética. El proceso de integración de ácido nucleico no nativo en el ADN genómico puede denominarse diseño genómico.

En algunos casos, el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que tenga, como máximo, un 10%, como máximo, un 15 %, como máximo, un 20 %, como máximo, un 30 %, como máximo, un 40 %, como máximo, un 50 %, como máximo, un 60 %, como máximo, un 70 %, como máximo, un 75 %, como máximo, un 80 %, como máximo, un 85 %, como máximo, un 90 %, como máximo, un 95 %, como máximo, un 99 % o un 100 % de identidad de secuencia de aminoácidos con un polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes,* SEQ ID NO: 8).

En algunos casos, el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que tenga al menos un 10 %, al menos un 15 %, 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 95 %, al menos un 99 % o 100 %, de identidad de secuencia de aminoácidos con un polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8).

En algunos casos, el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que tenga, como máximo, un 10%, como máximo, un 15 %, como máximo, un 20 %, como máximo, un 30 %, como máximo, un 40 %, como máximo, un 50 %, como máximo, un 60 %, como máximo, un 70 %, como máximo, un 75 %, como máximo, un 80 %, como máximo, un 85 %, como máximo, un 90 %, como máximo, un 95 %, como máximo, un 99 % o un 100 % de identidad de secuencia de aminoácidos con el dominio nucleasa de un polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes,* SEQ ID NO: 8).

Un polipéptido dirigido a un sitio puede comprender al menos un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con el polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos. Un polipéptido dirigido a un sitio puede comprender, como máximo, un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con el polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos. Un polipéptido dirigido a un sitio puede comprender al menos un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con un polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos en un dominio nucleasa HNH del polipéptido dirigido a un sitio. Un polipéptido dirigido a un sitio puede comprender, como máximo, un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con un polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos en un dominio nucleasa HNH del polipéptido dirigido a un sitio. Un polipéptido dirigido a un sitio puede comprender al menos un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con un polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos en un dominio nucleasa RuvC del polipéptido dirigido a un sitio. Un polipéptido dirigido a un sitio puede comprender, como máximo, un 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con un polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) sobre 10 aminoácidos contiguos en un dominio nucleasa RuvC del polipéptido dirigido a un sitio.

En algunos casos, el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que tenga al menos un 10 %, al menos un 15 %, al menos un 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al

menos un 60 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 95 %, al menos un 99 % o 100 % de identidad de secuencia de aminoácidos con el dominio nucleasa de un polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*).

El polipéptido dirigido a un sitio puede comprender una forma modificada de un polipéptido dirigido a un sitio ilustrativo de tipo silvestre. La forma modificada del polipéptido dirigido a un sitio ilustrativo de tipo silvestre puede comprender un cambio de aminoácido (por ejemplo, eliminación, inserción o sustitución) que reduzca la actividad de escisión de ácido nucleico del polipéptido dirigido a un sitio. Por ejemplo, la forma modificada del polipéptido dirigido a un sitio ilustrativo de tipo silvestre puede tener menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la actividad de escisión de ácido nucleico del polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*). La forma modificada del polipéptido dirigido a un sitio puede no tener actividad de escisión de ácido nucleico sustancial. Cuando un polipéptido dirigido a un sitio es una forma modificada que no tiene actividad de escisión de ácido nucleico sustancial, se puede denominar "inactivo enzimáticamente".

La forma modificada del polipéptido dirigido a un sitio ilustrativo de tipo silvestre puede tener más del 90 %, más del 80 %, más del 70 %, más del 60 %, más del 50 %, más del 40 %, más del 30 %, más del 20 %, más del 10 %, más del 5 % o más del 1 % de la actividad de escisión de ácido nucleico del polipéptido dirigido a un sitio ilustrativo de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*).

La forma modificada del polipéptido dirigido a un sitio puede comprender una mutación. La forma modificada del polipéptido dirigido a un sitio puede comprender una mutación que le permita inducir una rotura monocatenaria (SSB) en un ácido nucleico diana (por ejemplo, cortando solo una de las cadenas principales de azúcar-fosfato del ácido nucleico diana). La mutación puede dar lugar a menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la actividad de escisión de ácido nucleico en uno o más de la pluralidad de dominios de escisión de ácido nucleico del polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*). La mutación puede hacer que uno o más de la pluralidad de dominios de escisión de ácido nucleico conserve la capacidad de escindir la cadena complementaria del ácido nucleico diana, pero que tenga una menor capacidad de escindir la cadena no complementaria del ácido nucleico diana. La mutación puede hacer que uno o más de la pluralidad de dominios de escisión de ácido nucleico conserve la capacidad de escindir la cadena no complementaria del ácido nucleico diana, pero que tenga una menor capacidad de escindir la cadena complementaria del ácido nucleico diana. Por ejemplo, los restos del polipéptido Cas9 ilustrativo de tipo silvestre de *S. pyogenes* tales como Asp10, His840, Asn854 y Asn856 pueden mutarse para inactivar uno o más de la pluralidad de dominios de escisión de ácido nucleico (por ejemplo, dominios nucleasa). Los restos que se mutarán pueden corresponder a los restos Asp10, His840, Asn854 y Asn856 del polipéptido Cas9 ilustrativo de *S. pyogenes* de tipo silvestre (por ejemplo, según lo determinado por alineamiento de secuencia y/o estructural). Los ejemplos no limitantes de mutaciones pueden incluir D10A, H840A, N854A o N856A. Un experto en la materia reconocerá que son adecuadas las mutaciones distintas de las sustituciones de alanina.

Una mutación D10A se puede combinar con una o más de las mutaciones H840A, N854A o N856A para producir un polipéptido dirigido a un sitio que carezca sustancialmente de actividad de escisión de ADN. Una mutación H840A se puede combinar con una o más de las mutaciones D10A, N854A o N856A para producir un polipéptido dirigido a un sitio que carezca sustancialmente de actividad de escisión de ADN. Una mutación N854A se puede combinar con una o más de las mutaciones H840A, D10A o N856A para producir un polipéptido dirigido a un sitio que carezca sustancialmente de actividad de escisión de ADN. Una mutación N856A se puede combinar con una o más de las mutaciones H840A, N854A o D10A para producir un polipéptido dirigido a un sitio que carezca sustancialmente de actividad de escisión de ADN. Los polipéptidos dirigidos a un sitio que comprenden un dominio nucleasa sustancialmente inactivo pueden denominarse nickasas.

Las mutaciones de la divulgación se pueden producir por mutación dirigida a un sitio. Las mutaciones pueden incluir sustituciones, adiciones y eliminaciones, o cualquier combinación de las mismas. En algunos casos, la mutación convierte al aminoácido mutado en alanina. En algunos casos, la mutación convierte al aminoácido mutado en otro aminoácido (por ejemplo, glicina, serina, treonina, cisteína, valina, leucina, isoleucina, metionina, prolina, fenilalanina, tirosina, triptófano, ácido aspártico, ácido glutámico, asparaginas, glutamina, histidina, lisina o arginina). La mutación puede convertir al aminoácido mutado en un aminoácido no natural (por ejemplo, selenometionina). La mutación puede convertir al aminoácido mutado en miméticos de aminoácidos (por ejemplo, fosfomiméticos). La mutación puede ser una mutación conservativa. Por ejemplo, la mutación puede convertir al aminoácido mutado en aminoácidos que se parecen en tamaño, forma, carga, polaridad, configuración y/o rotámeros de los aminoácidos mutados (por ejemplo, mutación de cisteína/serina, mutación de lisina/asparagina, mutación de histidina/fenilalanina).

En algunos casos, el polipéptido dirigido a un sitio (por ejemplo, el polipéptido dirigido a un sitio variante, mutado, inactivo enzimáticamente y/o condicionalmente inactivo enzimáticamente) puede dirigirse al ácido nucleico. El polipéptido dirigido a un sitio (por ejemplo, endorribonucleasa variante, mutada, enzimáticamente inactiva y/o condicionalmente enzimáticamente inactiva) puede dirigirse a ARN. Los polipéptidos dirigidos a un sitio que pueden dirigirse al ARN pueden incluir miembros de otras subfamilias de CRISPR tales como Cas6 y Cas5.

El polipéptido dirigido a un sitio puede comprender una o más secuencias no nativas (por ejemplo, una fusión).

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), un dominio de unión a ácido nucleico y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC).

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*) y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC).

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*) y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de una bacteria (por ejemplo, *S. pyogenes*).

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que una al polipéptido dirigido a un sitio con una secuencia no nativa.

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %.

Un polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde uno de los dominios nucleasa comprende la mutación de ácido aspártico 10 y/o en donde uno de los dominios nucleasa comprende la mutación de histidina 840, y en donde la mutación reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %.

Endorribonucleasas

En algunas realizaciones, un polipéptido dirigido a un sitio puede ser una endorribonucleasa.

En algunos casos, la endorribonucleasa puede comprender una secuencia de aminoácidos que tenga, como máximo, aproximadamente un 20 %, como máximo, aproximadamente un 30 %, como máximo, aproximadamente un 40 %, como máximo, aproximadamente un 50 %, como máximo, aproximadamente un 60 %, como máximo, aproximadamente un 70 %, como máximo, aproximadamente un 75 %, como máximo, aproximadamente un 80 %, como máximo, aproximadamente un 85 %, como máximo, aproximadamente un 90 %, como máximo, aproximadamente un 95 %, como máximo, aproximadamente un 99 % o 100 % de identidad y/u homología de secuencia de aminoácidos con una endorribonucleasa de referencia de tipo silvestre. La endorribonucleasa puede comprender una secuencia de aminoácidos que tenga al menos aproximadamente un 20 %, al menos aproximadamente un 30 %, al menos aproximadamente un 40 %, al menos aproximadamente un 50 %, al menos aproximadamente un 60 %, al menos aproximadamente un 70 %, al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 99 % o un 100 % de identidad y/u homología de secuencia de aminoácidos con una endorribonucleasa de referencia de tipo silvestre (por ejemplo, Csy4 de *P. aeruginosa*). La endorribonucleasa de referencia puede ser un miembro de la familia Cas6 (por ejemplo, Csy4, Cas6). La endorribonucleasa de referencia puede ser un miembro de la familia Cas5 (por ejemplo, Cas5 de familia *D. vulgaris*). La endorribonucleasa de referencia puede ser un miembro de la familia CRISPR de Tipo I (por ejemplo, Cas3). Las endorribonucleasas de referencia pueden ser un miembro de la familia de Tipo II. La endorribonucleasa de referencia puede ser un miembro de la familia de Tipo III (por ejemplo, Cas6). Una endorribonucleasa de referencia puede ser miembro de la superfamilia de proteínas misteriosas asociadas con repeticiones (RAMP) (por ejemplo, Cas7).

La endorribonucleasa puede comprender modificaciones de aminoácidos (por ejemplo, sustituciones, eliminaciones, adiciones, etc.). La endorribonucleasa puede comprender una o más secuencias no nativas (por ejemplo, una fusión, un marcador de afinidad). Las modificaciones de aminoácidos pueden no alterar sustancialmente la actividad de la endorribonucleasa. Una endorribonucleasa que comprende modificaciones y/o fusiones de aminoácidos puede conservar al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente 97 % o 100 % de actividad de la endorribonucleasa de tipo silvestre.

La modificación puede producir la alteración de la actividad enzimática de la endorribonucleasa. La modificación puede dar lugar a menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %,

menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la endorribonucleasa. En algunos casos, la modificación se produce en el dominio nucleasa de una endorribonucleasa. Dichas modificaciones pueden dar lugar a menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la capacidad de escisión de ácido nucleico en uno o más de la pluralidad de dominios de escisión de ácido nucleico de la endorribonucleasa de tipo silvestre.

Endorribonucleasas condicionalmente inactivas enzimáticamente

En algunas realizaciones, una endorribonucleasa puede ser condicionalmente enzimáticamente inactiva. Una endorribonucleasa condicionalmente enzimáticamente inactiva puede unirse a un polinucleótido de una manera específica de la secuencia. Una endorribonucleasa condicionalmente enzimáticamente inactiva puede unirse a un polinucleótido de una manera específica de la secuencia, pero no puede escindir el polirribonucleótido diana.

En algunos casos, la endorribonucleasa condicionalmente enzimáticamente inactiva puede comprender una secuencia de aminoácidos que tenga hasta aproximadamente un 20 %, hasta aproximadamente un 30 %, hasta aproximadamente un 40 %, hasta aproximadamente un 50 %, hasta aproximadamente un 60 %, hasta aproximadamente un 70 %, hasta aproximadamente un 75 %, hasta aproximadamente un 80 %, hasta aproximadamente un 85 %, hasta aproximadamente un 90 %, hasta aproximadamente un 95 %, hasta aproximadamente un 99 % o un 100 % de identidad y/u homología de secuencia de aminoácidos con una endorribonucleasa condicionalmente enzimáticamente inactiva de referencia (por ejemplo, Csy4 de *P. aeruginosa*). En algunos casos, la endorribonucleasa condicionalmente enzimáticamente inactiva puede comprender una secuencia de aminoácidos que tenga al menos aproximadamente un 20 %, al menos aproximadamente un 30 %, al menos aproximadamente un 40 %, al menos aproximadamente un 50 %, al menos aproximadamente un 60 %, al menos aproximadamente un 70 %, al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 99 % o un 100 % de identidad y/u homología de secuencia de aminoácidos con una endorribonucleasa condicionalmente enzimáticamente inactiva de referencia (por ejemplo, Csy4 de *P. aeruginosa*).

La endorribonucleasa condicionalmente enzimáticamente inactiva puede comprender una forma modificada de una endorribonucleasa. La forma modificada de la endorribonucleasa puede comprender un cambio de aminoácido (por ejemplo, eliminación, inserción o sustitución) que reduzca la actividad de escisión de ácido nucleico de la endorribonucleasa. Por ejemplo, la forma modificada de la endorribonucleasa condicionalmente enzimáticamente inactiva puede tener menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la actividad de escisión de ácido nucleico de la endorribonucleasa condicionalmente enzimáticamente inactiva de referencia (por ejemplo, tipo silvestre) (por ejemplo, Csy4 de *P. aeruginosa*). La forma modificada de la endorribonucleasa condicionalmente enzimáticamente inactiva puede no tener actividad sustancial de escisión de ácido nucleico. Cuando una endorribonucleasa condicionalmente enzimáticamente inactiva es una forma modificada que no tiene una actividad sustancial de escisión de ácido nucleico, se puede denominar "inactiva enzimáticamente".

La forma modificada de la endorribonucleasa condicionalmente enzimáticamente inactiva puede comprender una mutación que puede dar lugar a una capacidad reducida de escisión de ácido nucleico (es decir, de modo que la endorribonucleasa condicionalmente enzimáticamente inactiva puede ser enzimáticamente inactiva en uno o más de los dominios de escisión de ácido nucleico). La mutación puede dar lugar a menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la capacidad de escisión de ácido nucleico en uno o más de la pluralidad de dominios de escisión de ácido nucleico de la endorribonucleasa de tipo silvestre (por ejemplo, Csy4 de *P. aeruginosa*). La mutación puede ocurrir en el dominio nucleasa de la endorribonucleasa. La mutación puede ocurrir en un pliegue de tipo ferredoxina. La mutación puede comprender la mutación de un aminoácido aromático conservado. La mutación puede comprender la mutación de un aminoácido catalítico. La mutación puede comprender la mutación de una histidina. Por ejemplo, la mutación puede comprender una mutación H29A en Csy4 (por ejemplo, Csy4 de *P. aeruginosa*) o cualquier resto correspondiente a H29A según lo determinado por la alineación de secuencia y/o estructural. Se pueden mutar otros restos para lograr el mismo efecto (es decir, inactivar uno o más de la pluralidad de dominios nucleasa).

Las mutaciones de la invención pueden producirse por mutación dirigida a un sitio. Las mutaciones pueden incluir sustituciones, adiciones y eliminaciones, o cualquier combinación de las mismas. En algunos casos, la mutación convierte al aminoácido mutado en alanina. En algunos casos, la mutación convierte al aminoácido mutado en otro aminoácido (por ejemplo, glicina, serina, treonina, cisteína, valina, leucina, isoleucina, metionina, prolina, fenilalanina, tirosina, triptófano, ácido aspártico, ácido glutámico, asparaginas, glutamina, histidina, lisina o arginina). La mutación puede convertir al aminoácido mutado en un aminoácido no natural (por ejemplo, selenometionina). La mutación puede convertir al aminoácido mutado en miméticos de aminoácidos (por ejemplo, fosfomiméticos). La mutación puede ser una mutación conservativa. Por ejemplo, la mutación puede convertir al aminoácido mutado en aminoácidos que se parecen en tamaño, forma, carga, polaridad, configuración y/o rotámeros de los aminoácidos mutados (por ejemplo, mutación de cisteína/serina, mutación de lisina/asparagina, mutación de histidina/fenilalanina).

Una endorribonucleasa condicionalmente enzimáticamente inactiva puede ser enzimáticamente inactiva en ausencia de un agente de reactivación (por ejemplo, imidazol). Un agente de reactivación puede ser un agente que imite a un resto de histidina (por ejemplo, puede tener un anillo de imidazol). Una endorribonucleasa condicionalmente enzimáticamente inactiva puede activarse mediante el contacto con un agente de reactivación. El agente de reactivación puede comprender imidazol. Por ejemplo, la endorribonucleasa condicionalmente enzimáticamente inactiva puede activarse enzimáticamente poniendo en contacto la endorribonucleasa condicionalmente enzimáticamente inactiva con imidazol a una concentración de aproximadamente 100 mM a aproximadamente 500 mM. El imidazol puede estar a una concentración de aproximadamente 100 mM, aproximadamente 150 mM, aproximadamente 200 mM, aproximadamente 250 mM, aproximadamente 300 mM, aproximadamente 350 mM, aproximadamente 400 mM, aproximadamente 450 mM, aproximadamente 500 mM, aproximadamente 550 mM o aproximadamente 600 mM. La presencia de imidazol (por ejemplo, en un intervalo de concentración de aproximadamente 100 mM a aproximadamente 500 mM) puede reactivar la endorribonucleasa condicionalmente enzimáticamente inactiva de modo que la endorribonucleasa condicionalmente enzimáticamente inactiva se vuelva enzimáticamente activa, por ejemplo, la endorribonucleasa condicionalmente enzimáticamente inactiva presente al menos aproximadamente un 50 %, al menos aproximadamente un 60 %, al menos aproximadamente un 70 %, al menos aproximadamente un 80 %, al menos aproximadamente un 90 %, al menos aproximadamente el 95 % o más de un 95 % de la capacidad de escisión de ácido nucleico de una endorribonucleasa condicionalmente enzimáticamente inactiva (por ejemplo, Csy4 de *P. aeruginosa* que comprende la mutación H29A).

Una endorribonucleasa condicionalmente enzimáticamente inactiva puede comprender al menos un 20 % de identidad de aminoácidos con Csy4 de *P. aeruginosa*, una mutación de histidina 29, en donde la mutación produce al menos un 50 % de reducción de la actividad de nucleasa de la endorribonucleasa, y en donde al menos el 50 % de la actividad de nucleasa perdida puede restablecerse mediante la incubación de la endorribonucleasa con al menos 100 mM de imidazol.

Optimización de codones

Un polinucleótido que codifica un polipéptido dirigido a un sitio y/o una endorribonucleasa puede someterse a una optimización de codones. Este tipo de optimización puede implicar la mutación del ADN extraído (por ejemplo, recombinante) para imitar las preferencias de codones del organismo o de la célula hospedadora previsto mientras codifica la misma proteína. Así pues, los codones se pueden cambiar, pero la proteína codificada permanece sin cambios. Por ejemplo, si la célula diana prevista fuera una célula humana, podría usarse un polinucleótido Cas9 con optimización de codones humano para producir un polipéptido dirigido a un sitio adecuado. Como otro ejemplo no limitante, si la célula hospedadora prevista fuera una célula de ratón, entonces un polinucleótido con optimización de codones de ratón que codifica Cas9 podría ser un polipéptido dirigido a un sitio adecuado. Un polinucleótido que codifica un polipéptido dirigido a un sitio puede estar optimizado en cuanto a los codones para muchas células hospedadoras de interés. Una célula hospedadora puede ser una célula de cualquier organismo (por ejemplo, una célula bacteriana, una célula de arquea, una célula de un organismo eucariota unicelular, una célula vegetal, una célula de algas, por ejemplo, *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens C. Agardh,* y similares, una célula fúngica (por ejemplo, una célula de levadura), una célula animal, una célula de un animal invertebrado (por ejemplo, mosca de la fruta, cnidario, equinodermo, nematodo, etc.), una célula de un animal vertebrado (por ejemplo, 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, etc.), etc. La optimización de codones puede no ser necesaria. En algunos casos, la optimización de codones puede ser preferible.

Ácido nucleico dirigido a ácido nucleico

La presente divulgación proporciona un ácido nucleico dirigido a ácido nucleico que puede dirigir las actividades de un polipéptido asociado (por ejemplo, un polipéptido dirigido a un sitio) a una secuencia diana específica dentro de un ácido nucleico diana. El ácido nucleico dirigido a ácido nucleico puede comprender nucleótidos. El ácido nucleico dirigido a ácido nucleico puede ser ARN. Un ácido nucleico dirigido a ácido nucleico puede comprender un ácido nucleico dirigido a ácido nucleico guía sencillo. Un ejemplo de ácido nucleico guía sencillo se representa en la **FIGURA 1A**. La extensión del espaciador **105** y la extensión del ARNcrtra **135** pueden comprender elementos que pueden aportar funcionalidad adicional (por ejemplo, estabilidad) al ácido nucleico dirigido a ácido nucleico. En algunas realizaciones, la extensión del espaciador **105** y la extensión de ARNcrtra **135** son opcionales. Una secuencia de espaciador **110** puede comprender una secuencia que puede hibridarse con una secuencia de ácido nucleico diana. La secuencia de espaciador **110** puede ser una parte variable del ácido nucleico dirigido a ácido nucleico. La secuencia de la secuencia de espaciador **110** puede diseñarse para hibridarse con la secuencia de ácido nucleico diana. La repetición CRISPR **115** (es decir, en la presente realización ilustrativa, como una repetición CRISPR mínima) puede comprender nucleótidos que pueden hibridarse con una secuencia de ARNcrtra **125** (es decir, en la presente realización ilustrativa, como una secuencia mínima de ARNcrtra). La repetición CRISPR mínima **115** y la secuencia mínima de ARNcrtra **125** pueden interactuar, comprendiendo las moléculas que interactúan una estructura bicatenaria de bases emparejadas. Conjuntamente, la repetición CRISPR mínima **115** y la secuencia mínima de ARNcrtra **125** pueden facilitar la unión al polipéptido dirigido a un sitio. La repetición mínima CRISPR **115** y la secuencia mínima de

ARNcrtra **125** pueden unirse para formar una estructura de horquilla a través del enlazador guía sencillo **120**. La secuencia de ARNcrtra **130** 3' puede comprender una secuencia de reconocimiento de motivos adyacentes protoespaciadores. La secuencia de ARNcrtra **130** 3' puede ser idéntica o similar a parte de una secuencia de ARNcrtra. En algunas realizaciones, la secuencia de ARNcrtra **130** 3' puede comprender una o más horquillas.

En algunas realizaciones, un ácido nucleico dirigido a ácido nucleico puede comprender un ácido nucleico dirigido a ácido nucleico guía sencillo como se representa en la **FIGURA 1B**. Un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de espaciador **140**. Una secuencia de espaciador **140** puede comprender una secuencia que puede hibridarse con la secuencia de ácido nucleico diana. La secuencia de espaciador **140** puede ser una parte variable del ácido nucleico dirigido a ácido nucleico. La secuencia de espaciador **140** puede estar a 5' de un primer dúplex **145**. El primer dúplex **145** comprende una región de hibridación entre una repetición CRISPR mínima **146** y una secuencia mínima de ARNcrtra **147**. El primer dúplex **145** puede ser interrumpido por una protuberancia **150**. La protuberancia **150** puede comprender nucleótidos no emparejados. La protuberancia **150** puede facilitar el reclutamiento de un polipéptido dirigido a un sitio hacia el ácido nucleico dirigido a ácido nucleico. La protuberancia **150** no está seguida de un primer lazo **155**. Una secuencia enlazadora une la repetición CRISPR mínima **146** y la secuencia mínima de ARNcrtra **147**. El último nucleótido emparejado en el extremo 3' del primer dúplex **145** se puede conectar a una segunda secuencia enlazadora **160**. El segundo enlazador **160** puede comprender un dominio P. El segundo enlazador **160** puede enlazar el primer dúplex **145** a un ARNcrtra medio **165**. El ARNcrtra medio **165** puede, en algunas realizaciones, comprender una o más regiones de horquilla. Por ejemplo, el ARNcrtra medio **165** puede comprender un segundo lazo **170** y un tercer lazo **180**. Un tercer enlazador **175** puede unir el segundo lazo **170** y el tercer lazo **180**.

En algunos aspectos, el ácido nucleico dirigido a ácido nucleico puede comprender una estructura de ácido nucleico guía doble. La **FIGURA 2** representa una estructura ilustrativa de ácido nucleico dirigido a ácido nucleico guía doble. Similar a la estructura de ácido nucleico guía sencillo de la **FIGURA 1**, la estructura de ácido nucleico guía doble puede comprender una extensión del espaciador **205**, un espaciador **210**, una repetición CRISPR mínima **215**, una secuencia mínima de ARNcrtra **230**, una secuencia de ARNcrtra 3' **235** y una extensión de ARNcrtra **240**. Sin embargo, un ácido nucleico dirigido a ácido nucleico guía doble puede no comprender el enlazador guía sencillo **120**. En cambio, la secuencia de repetición CRISPR mínima **215** puede comprender una secuencia de repetición CRISPR **220** 3' que puede ser similar o idéntica a parte de una repetición CRISPR. De igual manera, la secuencia mínima de ARNcrtra **230** puede comprender una secuencia de ARNcrtra 5' **225** que puede ser similar o idéntica a parte de un ARNcrtra. Los ARN guía dobles pueden hibridarse entre sí a través de la repetición CRISPR mínima **215** y la secuencia mínima de ARNcrtra **230**.

En algunas realizaciones, el primer segmento (es decir, el segmento dirigido a ácido nucleico) puede comprender la extensión del espaciador (por ejemplo, **105**/**205**) y el espaciador (por ejemplo, **110**/**210**). El ácido nucleico dirigido a ácido nucleico puede guiar al polipéptido unido a una secuencia de nucleótidos específica dentro del ácido nucleico diana a través del segmento dirigido a ácido nucleico mencionado anteriormente.

En algunas realizaciones, el segundo segmento (es decir, el segmento de unión a proteínas) puede comprender la repetición CRISPR mínima (por ejemplo, **115**/**215**), la secuencia mínima de ARNcrtra (por ejemplo, **125**/**230**), la secuencia de ARNcrtra 3' (por ejemplo, **130**/**235**) y/o la secuencia de extensión de ARNcrtra (por ejemplo, **135**/**240**). El segmento de unión a proteínas de un ácido nucleico dirigido a ácido nucleico puede interactuar con un polipéptido dirigido a un sitio. El segmento de unión a proteínas de un ácido nucleico dirigido a ácido nucleico puede comprender dos tramos de nucleótidos que pueden hibridarse entre sí. Los nucleótidos del segmento de unión a proteínas pueden hibridarse para formar un dúplex de ácido nucleico bicatenario. El dúplex de ácido nucleico bicatenario puede ser ARN. El dúplex de ácido nucleico bicatenario puede ser ADN.

En algunos casos, un ácido nucleico dirigido a ácido nucleico puede comprender, en el orden de 5' a 3', una extensión del espaciador, un espaciador, una repetición CRISPR mínima, un enlazador guía sencillo, un ARNcrtra mínimo, una secuencia de ARNcrtra 3' y una extensión de ARNcrtra. En algunos casos, un ácido nucleico dirigido a ácido nucleico puede comprender, una extensión de ARNcrtra, una secuencia de ARNcrtra 3', un ARNcrtra mínimo, un enlazador guía sencillo, una repetición CRISPR mínima, un espaciador y una extensión del espaciador en cualquier orden.

Un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio pueden formar un complejo. El ácido nucleico dirigido a ácido nucleico puede proporcionar especificidad por una diana al complejo al comprender una secuencia de nucleótidos que puede hibridarse con una secuencia de un ácido nucleico diana. En otras palabras, el polipéptido dirigido a un sitio puede guiarse a una secuencia de ácido nucleico en virtud de su asociación con al menos el segmento de unión a proteínas del ácido nucleico dirigido a ácido nucleico. El ácido nucleico dirigido a ácido nucleico puede dirigir la actividad de una proteína Cas9. El ácido nucleico dirigido a ácido nucleico puede dirigir la actividad de una proteína Cas9 enzimáticamente inactiva.

Los métodos de la divulgación pueden proporcionar una célula modificada genéticamente. Una célula modificada genéticamente puede comprender un ácido nucleico dirigido a ácido nucleico exógeno y/o un ácido nucleico exógeno que comprende una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico.

Secuencia de extensión del espaciador

Una secuencia de extensión del espaciador puede proporcionar estabilidad y/o proporcionar una ubicación para modificaciones de un ácido nucleico dirigido a ácido nucleico. Una secuencia de extensión del espaciador puede tener una longitud de aproximadamente 1 nucleótido a aproximadamente 400 nucleótidos. Una secuencia de extensión del espaciador puede tener una longitud de más de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 40,1000, 2000, 3000, 4000, 5000, 6000 o 7000, o más nucleótidos. Una secuencia de extensión del espaciador puede tener una longitud de menos de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000, 7000 o más nucleótidos. Una secuencia de extensión del espaciador puede tener menos de 10 nucleótidos de longitud. Una secuencia de extensión del espaciador puede tener entre 10 y 30 nucleótidos de longitud. Una secuencia de extensión del espaciador puede tener entre 30 y 70 nucleótidos de longitud.

La secuencia de extensión del espaciador puede comprender una fracción (por ejemplo, una secuencia de control de la estabilidad, una secuencia de unión a endorribonucleasa, una ribozima). Una fracción puede influir en la estabilidad de un ácido nucleico dirigido al ARN. Una fracción puede ser un segmento terminador de la transcripción (es decir, una secuencia de terminación de la transcripción). Una fracción de un ácido nucleico dirigido a ácido nucleico puede tener una longitud total de aproximadamente 10 nucleótidos a aproximadamente 100 nucleótidos, de aproximadamente 10 nucleótidos (nt) a aproximadamente 20 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 30 nt a aproximadamente 40 nt, de aproximadamente 40 nt a aproximadamente 50 nt, de aproximadamente 50 nt a aproximadamente 60 nt, de aproximadamente 60 nt a aproximadamente 70 nt, de aproximadamente 70 nt a aproximadamente 80 nt, de aproximadamente 80 nt a aproximadamente 90 nt, o de aproximadamente 90 nt a aproximadamente 100 nt, de aproximadamente 15 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 15 nt a aproximadamente 50 nt, de aproximadamente 15 nt a aproximadamente 40 nt, de aproximadamente 15 nt a aproximadamente 30 nt o de aproximadamente 15 nt a aproximadamente 25 nt. La fracción puede ser aquella que pueda funcionar en una célula eucariota. En algunos casos, la fracción puede ser aquella que pueda funcionar en una célula procariota. La fracción puede ser aquella que pueda funcionar tanto en una célula eucariota como en una célula procariota.

Los ejemplos no limitantes de fracciones adecuadas pueden incluir: caperuza de 5' (por ejemplo, una caperuza de 7-metilguanilato (m7 G)), una secuencia riboswitch (por ejemplo, para permitir una estabilidad regulada y/o accesibilidad regulada por proteínas y complejos de proteínas), una secuencia que forma un dúplex de ARNbc (es decir, una horquilla), una secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares), una modificación o secuencia que proporciona el seguimiento (por ejemplo, conjugación directa a una molécula fluorescente, conjugación a una fracción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.), una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluyendo activadores de la transcripción, represores de la transcripción, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares), una modificación o secuencia que proporciona una estabilidad aumentada, disminuida y/o controlable, o cualquier combinación de los mismos. Una secuencia de extensión del espaciador puede comprender un sitio de unión al cebador, un índice molecular (por ejemplo, secuencia de código de barras). La secuencia de extensión del espaciador puede comprender un marcador de afinidad de ácido nucleico.

Espaciador

El segmento de dirección a ácido nucleico de un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de nucleótidos (por ejemplo, un espaciador) que puede hibridarse con una secuencia de un ácido nucleico diana. El espaciador de un ácido nucleico dirigido a ácido nucleico puede interactuar con un ácido nucleico diana de una manera específica de la secuencia mediante hibridación (es decir, emparejamiento de bases). Por tanto, la secuencia de nucleótidos del espaciador puede variar y puede determinar la ubicación dentro del ácido nucleico diana en donde el ácido nucleico dirigido a ácido nucleico y el ácido nucleico diana pueden interactuar.

La secuencia de espaciador puede hibridarse con un ácido nucleico diana que se encuentra a 5' del motivo adyacente al espaciador (PAM). Diferentes organismos pueden comprender diferentes secuencias de PAM. Por ejemplo, en *S. pyogenes*, el PAM puede ser una secuencia del ácido nucleico diana que comprenda la secuencia 5'-XRR-3', donde R puede ser A o G, donde X es cualquier nucleótido y X está a inmediatamente 3' de la secuencia de ácido nucleico diana de la secuencia de espaciador.

La secuencia de ácido nucleico diana puede ser de 20 nucleótidos. El ácido nucleico diana puede tener menos de 20 nucleótidos. El ácido nucleico diana puede tener al menos 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o más nucleótidos. El ácido nucleico diana puede tener, como máximo, 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o más nucleótidos. La secuencia de ácido nucleico diana puede tener 20 bases a inmediatamente 5' del primer nucleótido del PAM. Por ejemplo, en una secuencia que comprende 5'-NNNNNNNNNNNNNNNNNNNNXRR-3', el ácido nucleico diana puede ser la secuencia que corresponde a los N, en donde N es cualquier nucleótido.

La secuencia dirigida a ácido nucleico del espaciador que puede hibridarse con el ácido nucleico diana puede tener una longitud de al menos aproximadamente 6 nt. Por ejemplo, la secuencia de espaciador que puede hibridarse con el ácido nucleico diana puede tener una longitud de al menos aproximadamente 6 nt, al menos aproximadamente 10 nt, al menos aproximadamente 15 nt, al menos aproximadamente 18 nt, al menos aproximadamente 19 nt, al menos aproximadamente 20 nt, al menos aproximadamente 25 nt, al menos aproximadamente 30 nt, al menos aproximadamente 35 nt o al menos aproximadamente 40 nt, de aproximadamente 6 nt a aproximadamente 80 nt, de aproximadamente 6 nt a aproximadamente 50nt, de aproximadamente 6 nt a aproximadamente 45 nt, de aproximadamente 6 nt a aproximadamente 40 nt, de aproximadamente 6 nt a aproximadamente 35 nt, de aproximadamente 6 nt a aproximadamente 30 nt, de aproximadamente 6 nt a aproximadamente 25 nt, de aproximadamente 6 nt a aproximadamente 20 nt, de aproximadamente 6 nt a aproximadamente 19 nt, de aproximadamente 10 nt a aproximadamente 50nt, de aproximadamente 10 nt a aproximadamente 45 nt, de aproximadamente 10 nt a aproximadamente 40 nt, de aproximadamente 10 nt a aproximadamente 35 nt, de aproximadamente 10 nt a aproximadamente 30 nt, de aproximadamente 10 nt a aproximadamente 25 nt, de aproximadamente 10 nt a aproximadamente 20 nt, de aproximadamente 10 nt a aproximadamente 19 nt, de aproximadamente 19 nt a aproximadamente 25 nt, de aproximadamente 19 nt a aproximadamente 30 nt, de aproximadamente 19 nt a aproximadamente 35 nt, de aproximadamente 19 nt a aproximadamente 40 nt, de aproximadamente 19 nt a aproximadamente 45 nt, de aproximadamente 19 nt a aproximadamente 50 nt, de aproximadamente 19 nt a aproximadamente 60 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 20 nt a aproximadamente 35 nt, de aproximadamente 20 nt a aproximadamente 40 nt, de aproximadamente 20 nt a aproximadamente 45 nt, de aproximadamente 20 nt a aproximadamente 50 nt o de aproximadamente 20 nt a aproximadamente 60 nt. En algunos casos, la secuencia de espaciador que puede hibridarse con el ácido nucleico diana puede tener 20 nucleótidos de longitud. El espaciador que puede hibridarse con el ácido nucleico diana puede tener 19 nucleótidos de longitud.

El porcentaje de complementariedad entre la secuencia de espaciador del ácido nucleico diana puede ser de al menos aproximadamente 30 %, al menos aproximadamente 40 %, al menos aproximadamente 50 %, al menos aproximadamente 60 %, al menos aproximadamente 65 %, al menos aproximadamente 70 %, al menos aproximadamente 75 %, al menos aproximadamente 80 %, al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o 100 %. El porcentaje de complementariedad entre la secuencia de espaciador del ácido nucleico diana puede ser, como máximo, de aproximadamente 30 %, como máximo, aproximadamente 40 %, como máximo, aproximadamente 50 %, como máximo, aproximadamente 60 %, como máximo, aproximadamente 65 %, como máximo, aproximadamente 70 %, como máximo, aproximadamente 75 %, como máximo, aproximadamente 80 %, como máximo, aproximadamente 85 %, como máximo, aproximadamente 90 %, como máximo, aproximadamente 95 %, como máximo, aproximadamente 97 %, como máximo, aproximadamente 98 %, como máximo, aproximadamente 99 % o 100 %. En algunos casos, el porcentaje de complementariedad entre la secuencia de espaciador y el ácido nucleico diana puede ser del 100 % sobre los seis nucleótidos contiguos más a 5' de la secuencia diana de la cadena complementaria del ácido nucleico diana. En algunos casos, el porcentaje de complementariedad entre la secuencia de espaciador y el ácido nucleico diana puede ser al menos del 60 % sobre aproximadamente 20 nucleótidos contiguos. En algunos casos, el porcentaje de complementariedad entre la secuencia de espaciador y el ácido nucleico diana puede ser del 100 % sobre los catorce nucleótidos contiguos más a 5' de la secuencia diana de la cadena complementaria del ácido nucleico diana y tan bajo como del 0 % en el resto. En dicho caso, se puede considerar que la secuencia de espaciador tiene 14 nucleótidos de longitud. En algunos casos, el porcentaje de complementariedad entre la secuencia de espaciador y el ácido nucleico diana puede ser del 100 % sobre los seis nucleótidos contiguos más a 5' de la secuencia diana de la cadena complementaria del ácido nucleico diana y tan bajo como del 0 % en el resto. En dicho caso, se puede considerar que la secuencia de espaciador 6 nucleótidos de longitud. El ácido nucleico diana puede ser más de aproximadamente el 50 %, 60 %, 70 %, 80 %, 90 % o 100 % complementario a la región semilla del ARNcr. El ácido nucleico diana puede ser menos del aproximadamente 50 %, 60 %, 70 %, 80 %, 90 % o 100 % complementario a la región semilla del ARNcr.

El segmento espaciador de un ácido nucleico dirigido a ácido nucleico puede modificarse (por ejemplo, mediante ingeniería genética) para que se hibride con cualquier secuencia deseada dentro de un ácido nucleico diana. Por ejemplo, se puede diseñar un espaciador (por ejemplo, diseñar, programar) para que se hibride con una secuencia en ácido nucleico diana que participe en el cáncer, el crecimiento celular, la replicación del ADN, la reparación del ADN, genes HLA, proteínas de la superficie celular, receptores de linfocitos T, genes de la superfamilia de inmunoglobulinas, genes supresores de tumores, genes de microARN, genes de ARN no codificantes largos, factores de transcripción, globinas, proteínas víricas, genes mitocondriales y similares.

Se puede identificar una secuencia de espaciador usando un programa informático (por ejemplo, código legible por máquina). El programa informático puede usar variables tales como la temperatura de fusión prevista, formación de estructuras secundarias y temperatura de hibridación previstas, identidad de secuencia, contexto genómico, accesibilidad a la cromatina, % de GC, frecuencia de aparición genómica, estado de metilación, presencia de SNP y similares.

Secuencia de repeticiones CRISPR mínima

Una secuencia de repetición CRISPR mínima puede ser una secuencia de al menos aproximadamente un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/o homología de secuencia con una secuencia de repetición CRISPR de referencia (por ejemplo, ARNcr de *S. pyogenes*). Una secuencia de repetición CRISPR mínima puede ser una secuencia con, como máximo, un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/u homología de secuencia con una secuencia de repetición CRISPR de referencia (por ejemplo, ARNcr de *S. pyogenes*). Una repetición CRISPR mínima puede comprender nucleótidos que pueden hibridarse con una secuencia mínima de ARNcrtra. Una repetición CRISPR mínima y una secuencia mínima de ARNcrtra pueden formar una estructura bicatenaria de bases emparejadas. Conjuntamente, la repetición CRISPR mínima y la secuencia mínima de ARNcrtra pueden facilitar la unión al polipéptido dirigido a un sitio. Una parte de la secuencia de repetición CRISPR mínima puede hibridarse con la secuencia mínima de ARNcrtra. Una parte de la secuencia de repetición CRISPR mínima puede ser al menos aproximadamente un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % complementaria a la secuencia mínima de ARNcrtra. Una parte de la secuencia de repetición CRISPR mínima puede ser, como máximo, un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % complementaria a la secuencia mínima de ARNcrtra.

Una secuencia de repetición CRISPR mínima puede tener una longitud de aproximadamente 6 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, la secuencia de repetición CRISPR mínima puede tener una longitud de aproximadamente 6 nucleótidos (nt) a aproximadamente 50 nt, de aproximadamente 6 nt a aproximadamente 40 nt, de aproximadamente 6 nt a aproximadamente 30 nt, de aproximadamente 6 nt a aproximadamente 25 nt, de aproximadamente 6 nt a aproximadamente 20 nt, de aproximadamente 6 nt a aproximadamente 15 nt, de aproximadamente 8 nt a aproximadamente 40 nt, de aproximadamente 8 nt a aproximadamente 30 nt, de aproximadamente 8 nt a aproximadamente 25 nt, de aproximadamente 8 nt a aproximadamente 20 nt o de aproximadamente 8 nt a aproximadamente 15 nt, de aproximadamente 15 nt a aproximadamente 100 nt, de aproximadamente 15 nt a aproximadamente 80 nt, de aproximadamente 15 nt a aproximadamente 50 nt, de aproximadamente 15 nt a aproximadamente 40 nt, de aproximadamente 15 nt a aproximadamente 30 nt o de aproximadamente 15 nt a aproximadamente 25 nt. En algunas realizaciones, la secuencia de repetición CRISPR mínima tiene una longitud de aproximadamente 12 nucleótidos.

La secuencia de repetición CRISPR mínima puede ser al menos aproximadamente un 60 % idéntica a una secuencia de repetición CRISPR mínima de referencia (por ejemplo, ARNc de tipo silvestre de *S. pyogenes*) en un tramo de al menos 6, 7 u 8 nucleótidos contiguos. La secuencia de repetición CRISPR mínima puede ser al menos aproximadamente un 60 % idéntica a una secuencia de repetición CRISPR mínima de referencia (por ejemplo, ARNc de tipo silvestre de *S. pyogenes*) en un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, La secuencia de repetición CRISPR mínima puede ser al menos aproximadamente un 65 % idéntica, al menos aproximadamente un 70 % idéntica, al menos aproximadamente un 75 % idéntica, al menos aproximadamente un 80 % idéntica, al menos aproximadamente un 85 % idéntica, al menos aproximadamente un 90 % idéntico, al menos aproximadamente un 95 % idéntica, al menos aproximadamente un 98 % idéntica, al menos aproximadamente un 99 % idéntica o un 100 % idéntica a una secuencia de repetición CRISPR mínima de referencia en un tramo de al menos 6, 7 u 8 nucleótidos contiguos.

Secuencia mínima de ARNcrtra

Una secuencia mínima de ARNcrtra puede ser una secuencia con al menos aproximadamente un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/o homología de secuencia con una secuencia de ARNcrtra de referencia (por ejemplo, ARNcrtra de tipo silvestre de *S. pyogenes*). Una secuencia mínima de ARNcrtra puede ser una secuencia con, como máximo, aproximadamente un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/o homología de secuencia con una secuencia de ARNcrtra de referencia (por ejemplo, ARNcrtra de tipo silvestre de *S. pyogenes*). Una secuencia mínima de ARNcrtra puede comprender nucleótidos que se pueden hibridar con una secuencia de repetición de CRISPR mínima. Una secuencia mínima de ARNcrtra y una secuencia de repetición CRISPR mínima pueden formar una estructura bicatenaria de bases emparejadas. Conjuntamente, la secuencia mínima de ARNcrtra y la secuencia de repetición CRISPR mínima pueden facilitar la unión al polipéptido dirigido a un sitio. Una parte de la secuencia mínima de ARNcrtra se puede hibridar con la secuencia de repetición CRISPR mínima. Una parte de la secuencia mínima de ARNcrtra puede ser un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % complementaria a la secuencia de repetición CRISPR mínima.

La secuencia mínima de ARNcrtra puede tener una longitud de aproximadamente 6 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, la secuencia mínima de ARNcrtra puede tener una longitud de aproximadamente 6 nucleótidos (nt) a aproximadamente 50 nt, de aproximadamente 6 nt a aproximadamente 40 nt, de aproximadamente 6 nt a aproximadamente 30 nt, de aproximadamente 6 nt a aproximadamente 25 nt, de aproximadamente 6 nt a aproximadamente 20 nt, de aproximadamente 6 nt a aproximadamente 15 nt, de aproximadamente 8 nt a aproximadamente 40 nt, de aproximadamente 8 nt a aproximadamente 30 nt, de aproximadamente 8 nt a aproximadamente 25 nt, de aproximadamente 8 nt a aproximadamente 20 nt o de aproximadamente 8 nt a aproximadamente 15 nt, de aproximadamente 15 nt a aproximadamente 100 nt, de aproximadamente 15 nt a

aproximadamente 80 nt, de aproximadamente 15 nt a aproximadamente 50 nt, de aproximadamente 15 nt a aproximadamente 40 nt, de aproximadamente 15 nt a aproximadamente 30 nt o de aproximadamente 15 nt a aproximadamente 25 nt. En algunas realizaciones, la secuencia mínima de ARNcrtra tiene una longitud de aproximadamente 14 nucleótidos.

La secuencia mínima de ARNcrtra puede ser al menos aproximadamente un 60 % idéntica a una secuencia mínima de ARNcrtra de referencia (por ejemplo, ARNcrtra de tipo silvestre de *S. pyogenes*) en un tramo de al menos 6, 7 u 8 nucleótidos contiguos. La secuencia mínima de ARNcrtra puede ser al menos aproximadamente un 60 % idéntica a una secuencia mínima de ARNcrtra de referencia (por ejemplo, ARNcrtra de tipo silvestre de *S. pyogenes*) en un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, la secuencia mínima de ARNcrtra puede ser al menos aproximadamente un 65 % idéntica, al menos aproximadamente un 70 % idéntica, al menos aproximadamente un 75 % idéntica, al menos aproximadamente un 80 % idéntica, al menos aproximadamente un 85 % idéntica, al menos aproximadamente un 90 % idéntico, al menos aproximadamente un 95 % idéntica, al menos aproximadamente un 98 % idéntica, al menos aproximadamente 99 % idéntica o 100 % idéntica a una secuencia mínima de ARNcrtra de referencia en un tramo de al menos 6, 7 u 8 nucleótidos contiguos.

El dúplex (es decir, el primer dúplex de la **Figura 1B**) entre el ARN de CRISPR mínimo y el ARNcrtra mínimo puede comprender una doble hélice. La primera base de la primera cadena del dúplex (por ejemplo, la repetición CRISPR mínima de la **Figura 1B**) puede ser una guanina. La primera base de la primera cadena del dúplex (por ejemplo, la repetición CRISPR mínima de la **Figura 1B**) puede ser una adenina. El dúplex (es decir, el primer dúplex de la **Figura 1B**) entre el ARN de CRISPR mínimo y el ARNcrtra mínimo puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más nucleótidos. El dúplex (es decir, el primer dúplex de la **Figura 1B**) entre el ARN de CRISPR mínimo y el ARNcrtra mínimo puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos.

El dúplex puede comprender una falta de coincidencia. El dúplex puede comprender al menos aproximadamente 1, 2, 3, 4, 5 o más faltas de coincidencia. El dúplex puede comprender, como máximo, aproximadamente 1, 2, 3, 4 o 5 faltas de coincidencia. En algunos casos, el dúplex comprende no más de 2 faltas de coincidencia.

Protuberancia

Una protuberancia puede referirse a una región no emparejada de nucleótidos dentro del dúplex formado por la repetición CRISPR mínima y la secuencia mínima de ARNcrtra. La protuberancia puede ser importante en la unión al polipéptido dirigido a un sitio. una protuberancia puede comprender, a un lado del dúplex, un 5'-XXXY-3' no emparejado, donde X es cualquier purina e Y puede ser un nucleótido que puede formar un par oscilante con un nucleótido en la cadena opuesta, y una región de nucleótidos no emparejados en el otro lado del dúplex.

Por ejemplo, la protuberancia puede comprender una purina no emparejada (por ejemplo, adenina) en la cadena de repetición CRISPR mínima de la protuberancia. En algunas realizaciones, una protuberancia puede comprender un 5'-AAGY-3' no emparejado de la cadena de secuencia de ARNcrtra mínima de la protuberancia, donde Y puede ser un nucleótido que puede formar un emparejamiento oscilante con un nucleótido en la cadena de repetición CRISPR mínima.

Una protuberancia en un primer lado del dúplex (por ejemplo, el lado de la repetición CRISPR mínima) puede comprender al menos 1, 2, 3, 4 o 5, o más nucleótidos no emparejados. Una protuberancia en un primer lado del dúplex (por ejemplo, el lado de la repetición CRISPR mínima) puede comprender, como máximo, 1, 2, 3, 4 o 5, o más nucleótidos no emparejados. Una protuberancia en el primer lado del dúplex (por ejemplo, el lado de la repetición CRISPR mínima) puede comprender 1 nucleótido no emparejado.

Una protuberancia en un segundo lado del dúplex (por ejemplo, el lado de la secuencia mínima de ARNcrtra del dúplex) puede comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más nucleótidos no emparejados. Una protuberancia en un segundo lado del dúplex (por ejemplo, el lado de la secuencia mínima de ARNcrtra del dúplex) puede comprender, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más nucleótidos no emparejados. Una protuberancia en un segundo lado del dúplex (por ejemplo, el lado de la secuencia mínima de ARNcrtra del dúplex) puede comprender 4 nucleótidos no emparejados.

Las regiones de diferentes números de nucleótidos no emparejados en cada cadena del dúplex se pueden emparejar entre sí. Por ejemplo, una protuberancia puede comprender 5 nucleótidos no emparejados de una primera cadena y 1 nucleótido no emparejado de una segunda cadena. Una protuberancia puede comprender 4 nucleótidos no emparejados de una primera cadena y 1 nucleótido no emparejado de una segunda cadena. Una protuberancia puede comprender 3 nucleótidos no emparejados de una primera cadena y 1 nucleótido no emparejado de una segunda cadena. Una protuberancia puede comprender 2 nucleótidos no emparejados de una primera cadena y 1 nucleótido no emparejado de una segunda cadena. Una protuberancia puede comprender 1 nucleótido no emparejado de una primera cadena y 1 nucleótido no emparejado de una segunda cadena. Una protuberancia puede comprender 1 nucleótido no emparejado de una primera cadena y 2 nucleótidos no emparejados de una segunda cadena. Una protuberancia puede comprender 1 nucleótido no emparejado de una primera cadena y 3 nucleótido no emparejado

de una segunda cadena. Una protuberancia puede comprender 1 nucleótido no emparejado de una primera cadena y 4 nucleótidos no emparejados de una segunda cadena. una protuberancia puede comprender 1 nucleótido no emparejado de una primera cadena y 5 nucleótidos no emparejados de una segunda cadena.

En algunos casos, una protuberancia puede comprender al menos un emparejamiento oscilante. En algunos casos, una protuberancia puede comprender, como máximo, un emparejamiento oscilante. Una secuencia de protuberancia puede comprender al menos un nucleótido de purina. Una secuencia de protuberancia puede comprender al menos 3 nucleótidos de purina. Una secuencia de protuberancia puede comprender al menos 5 nucleótidos de purina. Una secuencia de protuberancia puede comprender al menos un nucleótido de guanina. Una secuencia de protuberancia puede comprender al menos un nucleótido de adenina.

Dominio P (DOMINIO P)

Un dominio P puede referirse a una región de un ácido nucleico dirigido a un ácido nucleico que puede reconocer un motivo adyacente al protoespaciador (PAM) en un ácido nucleico diana. Un dominio P puede hibridarse con un PAM en un ácido nucleico diana. Por tanto, un dominio P puede comprender una secuencia que sea complementaria a un PAM. Un dominio P puede ubicarse a 3' de la secuencia mínima de ARNcrtra. Un dominio P puede ubicarse dentro de una secuencia de ARNcrtra de 3' (es decir, una secuencia de ARNcrtra media).

Un dominio P comienza al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20, o más nucleótidos 3' del último nucleótido emparejado del dúplex de repetición CRISPR mínima y secuencia mínima de ARNcrtra. Un dominio P puede comenzar, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más nucleótidos a 3' del último nucleótido emparejado del dúplex de repetición CRISPR mínima y secuencia mínima de ARNcrtra.

Un dominio P puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20, o más nucleótidos consecutivos. Un dominio P puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20, o más nucleótidos consecutivos.

En algunos casos, un dominio P puede comprender un dinucleótido CC (es decir, dos nucleótidos de citosina consecutivos). El dinucleótido CC puede interactuar con el dinucleótido GG de un PAM, en donde el PAM comprende una secuencia 5'-XGG-3'.

Un dominio P puede ser una secuencia de nucleótidos ubicada en la secuencia de ARNcrtra 3' (es decir, la secuencia de ARNcrtra media). Un dominio P puede comprender nucleótidos duplicados (por ejemplo, nucleótidos de una horquilla, hibridados entre sí. Por ejemplo, un dominio P puede comprender un dinucleótido CC que se hibrida con un dinucleótido GG en un dúplex en horquilla de la secuencia de ARNcrtra 3' (es decir, la secuencia de ARNcrtra media). La actividad del dominio P (por ejemplo, la capacidad del ácido nucleico dirigido a ácido nucleico para dirigirse a un ácido nucleico diana) puede estar regulada por el estado de hibridación del DOMINIO P. Por ejemplo, si el dominio P se hibrida, el ácido nucleico dirigido a ácido nucleico puede no reconocer su diana. Si el dominio P no se hibrida, el ácido nucleico dirigido a ácido nucleico puede reconocer su diana.

El dominio P puede interactuar con regiones que interactúan con el dominio P dentro del polipéptido dirigido a un sitio. El dominio P puede interactuar con un parche básico rico en arginina en el polipéptido dirigido a un sitio. Las regiones que interactúan con el dominio P pueden interactuar con una secuencia de PAM. El dominio P puede comprender un tallo-lazo. El dominio P puede comprender una protuberancia.

Secuencia de ARNcrtra de 3'

Una secuencia de ARNcrtra de 3' puede ser una secuencia con al menos aproximadamente un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/u homología de secuencia con una secuencia de ARNcrtra de referencia (por ejemplo, ARNcrtra de *S. pyogenes*). Una secuencia de ARNcrtra de 3' puede ser una secuencia con, como máximo, un 30 %, 40 %, 50 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia y/u homología de secuencia con una secuencia de ARNcrtra de referencia (por ejemplo, ARNcrtra de *S. pyogenes*).

La secuencia de ARNcrtra de 3' puede tener una longitud de aproximadamente 6 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, la secuencia de ARNcrtra de 3' puede tener una longitud de aproximadamente 6 nucleótidos (nt) a aproximadamente 50 nt, de aproximadamente 6 nt a aproximadamente 40 nt, de aproximadamente 6 nt a aproximadamente 30 nt, de aproximadamente 6 nt a aproximadamente 25 nt, de aproximadamente 6 nt a aproximadamente 20 nt, de aproximadamente 6 nt a aproximadamente 15 nt, de aproximadamente 8 nt a aproximadamente 40 nt, de aproximadamente 8 nt a aproximadamente 30 nt, de aproximadamente 8 nt a aproximadamente 25 nt, de aproximadamente 8 nt a aproximadamente 20 nt o de aproximadamente 8 nt a aproximadamente 15 nt, de aproximadamente 15 nt a aproximadamente 100 nt, de aproximadamente 15 nt a aproximadamente 80 nt, de aproximadamente 15 nt a aproximadamente 50 nt, de aproximadamente 15 nt a aproximadamente 40 nt, de aproximadamente 15 nt a aproximadamente 30 nt o de aproximadamente 15 nt a aproximadamente 25 nt. En algunas realizaciones, la secuencia de ARNcrtra de 3' tiene una longitud de

aproximadamente 14 nucleótidos.

La secuencia de ARNcrtra de 3' puede ser al menos aproximadamente un 60 % idéntica a una secuencia de ARNcrtra de 3' de referencia (por ejemplo, la secuencia de ARNcrtra de 3' de tipo silvestre de *S. pyogenes*) sobre un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, la secuencia de ARNcrtra de 3' puede ser al menos aproximadamente un 60 % idéntica, al menos aproximadamente un 65 % idéntica, al menos aproximadamente un 70 % idéntica, al menos aproximadamente un 75 % idéntica, al menos aproximadamente un 80 % idéntica, al menos aproximadamente un 85 % idéntica, al menos aproximadamente un 90 % idéntico, al menos aproximadamente un 95 % idéntica, al menos aproximadamente un 98 % idéntica, al menos aproximadamente un 99 % idéntica o un 100 % idéntica, a una secuencia de ARNcrtra de 3' de referencia (por ejemplo, secuencia de ARNcrtra de 3' de tipo silvestre de *S. pyogenes*) en un tramo de al menos 6, 7 u 8 nucleótidos contiguos.

Una secuencia de ARNcrtra de 3' puede comprender más de una región duplexada (por ejemplo, horquilla, región hibridada). Una secuencia de ARNcrtra de 3' puede comprender dos regiones duplexadas.

La secuencia de ARNcrtra de 3' también puede denominarse ARNcrtra media (véase la **Figura 1B**). La secuencia de ARNcrtra media puede comprender una estructura de tallo-lazo. En otras palabras, la secuencia del ARNcrtra media puede comprender una horquilla que sea diferente de un segundo o tercer tallo, como se muestra en la **Figura 1B**. Una estructura de tallo-lazo de ARNcrtra medio (es decir, el ARNcrtra de 3') puede comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20, o más nucleótidos. Una estructura de tallo-lazo en el ARNcrtra medio (es decir, el ARNcrtra de 3') puede comprender, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más nucleótidos. La estructura de tallo-lazo puede comprender una fracción funcional. Por ejemplo, la estructura de tallo-lazo puede comprender un aptámero, una ribozima, una horquilla que interactúa con proteínas, una matriz CRISPR, un intrón y un exón. La estructura de tallo-lazo puede comprender al menos aproximadamente 1, 2, 3, 4 o 5 fracciones funcionales. La estructura de tallo-lazo puede comprender, como máximo, aproximadamente 1, 2, 3, 4 o 5, o más fracciones funcionales.

La horquilla de la secuencia de ARNcrtra medio puede comprender un dominio P. El dominio P puede comprender una región bicatenaria en la horquilla.

Secuencia de extensión de ARNcrtra

Una secuencia de extensión del ARNcrtra puede proporcionar estabilidad y/o proporcionar una ubicación para modificaciones de un ácido nucleico dirigido a ácido nucleico. Una secuencia de extensión del ARNcrtra puede tener una longitud de aproximadamente 1 nucleótido a aproximadamente 400 nucleótidos. Una secuencia de extensión del ARNcrtra puede tener una longitud de más de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400 o más nucleótidos. Una secuencia de extensión del ARNcrtra puede tener una longitud de aproximadamente 20 a aproximadamente 5000 o más nucleótidos. Una secuencia de extensión del ARNcrtra puede tener una longitud de más de 1000 nucleótidos. Una secuencia de extensión del ARNcrtra puede tener una longitud de menos de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400 nucleótidos. Una secuencia de extensión del ARNcrtra puede tener una longitud de menos de 1000 nucleótidos. Una secuencia de extensión del ARNcrtra puede tener menos de 10 nucleótidos de longitud. Una secuencia de extensión del ARNcrtra puede tener entre 10 y 30 nucleótidos de longitud. Una secuencia de extensión del ARNcrtra puede tener entre 30 y 70 nucleótidos de longitud.

La secuencia de extensión del ARNcrtra puede comprender una fracción (por ejemplo, secuencia de control de la estabilidad, ribozima, secuencia de unión a endorribonucleasa). Una fracción puede influir en la estabilidad de un ácido nucleico dirigido al ARN. Una fracción puede ser un segmento terminador de la transcripción (es decir, una secuencia de terminación de la transcripción). Una fracción de un ácido nucleico dirigido a ácido nucleico puede tener una longitud total de aproximadamente 10 nucleótidos a aproximadamente 100 nucleótidos, de aproximadamente 10 nucleótidos (nt) a aproximadamente 20 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 30 nt a aproximadamente 40 nt, de aproximadamente 40 nt a aproximadamente 50 nt, de aproximadamente 50 nt a aproximadamente 60 nt, de aproximadamente 60 nt a aproximadamente 70 nt, de aproximadamente 70 nt a aproximadamente 80 nt, de aproximadamente 80 nt a aproximadamente 90 nt, o de aproximadamente 90 nt a aproximadamente 100 nt, de aproximadamente 15 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 15 nt a aproximadamente 50 nt, de aproximadamente 15 nt a aproximadamente 40 nt, de aproximadamente 15 nt a aproximadamente 30 nt o de aproximadamente 15 nt a aproximadamente 25 nt. La fracción puede ser aquella que pueda funcionar en una célula eucariota. En algunos casos, la fracción puede ser aquella que pueda funcionar en una célula procariota. La fracción puede ser aquella que pueda funcionar tanto en una célula eucariota como en una célula procariota.

Los ejemplos no limitantes de fracciones de extensión de ARNcrtra adecuadas incluyen: una cola 3’ poli-adenilada, una secuencia riboswitch (por ejemplo, para permitir una estabilidad regulada y/o accesibilidad regulada por proteínas y complejos de proteínas), una secuencia que forma un dúplex de ARNbc (es decir, una horquilla), una secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares), una modificación o secuencia que proporciona el seguimiento (por ejemplo, conjugación directa a una molécula fluorescente,

conjugación a una fracción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.), una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluyendo activadores de la transcripción, represores de la transcripción, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares), una modificación o secuencia que proporciona una estabilidad aumentada, disminuida y/o controlable, o cualquier combinación de los mismos. Una secuencia de extensión del ARNcrtra puede comprender un sitio de unión al cebador, un índice molecular (por ejemplo, secuencia de código de barras). En algunas realizaciones de la divulgación, la secuencia de extensión del ARNcrtra puede comprender uno o más marcadores de afinidad.

Ácido nucleico guía sencillo

El ácido nucleico dirigido a ácido nucleico puede ser un ácido nucleico guía sencillo. El ácido nucleico guía sencillo puede ser ARN. Un ácido nucleico guía sencillo puede comprender un enlazador (es decir, el elemento **120** de la **Figura 1A**) entre la secuencia de repetición CRISPR mínima y la secuencia de ARNcrtra mínima, que puede denominarse secuencia de enlazador guía sencillo.

El enlazador guía sencillo de un ácido nucleico guía sencillo puede tener una longitud de aproximadamente 3 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, el enlazador puede tener una longitud de aproximadamente 3 nucleótidos (nt) a aproximadamente 90 nt, de aproximadamente 3 nt a aproximadamente 80 nt, de aproximadamente 3 nt a aproximadamente 70 nt, de aproximadamente 3 nt a aproximadamente 60 nt, de aproximadamente 3 nt a aproximadamente 50 nt, de aproximadamente 3 nt a aproximadamente 40 nt, de aproximadamente 3 nt a aproximadamente 30 nt, de aproximadamente 3 nt a aproximadamente 20 nt o de aproximadamente 3 nt a aproximadamente 10 nt. Por ejemplo, El enlazador puede tener una longitud de aproximadamente 3 nt a aproximadamente 5 nt, de aproximadamente 5 nt a aproximadamente 10 nt, de aproximadamente 10 nt a aproximadamente 15 nt, de aproximadamente 15 nt a aproximadamente 20 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 25 nt a aproximadamente 30 nt, de aproximadamente 30 nt a aproximadamente 35 nt, de aproximadamente 35 nt a aproximadamente 40 nt, de aproximadamente 40 nt a aproximadamente 50 nt, de aproximadamente 50 nt a aproximadamente 60 nt, de aproximadamente 60 nt a aproximadamente 70 nt, de aproximadamente 70 nt a aproximadamente 80 nt, de aproximadamente 80 nt a aproximadamente 90 nt. de aproximadamente 90 nt a aproximadamente 100 nt. En algunas realizaciones, El enlazador de un ácido nucleico guía sencillo es de entre 4 y 40 nucleótidos. Un enlazador puede tener una longitud de al menos aproximadamente 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 o 7000, o más nucleótidos. Un enlazador puede tener una longitud, como máximo, de 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 o 7000, o más nucleótidos.

La secuencia enlazadora puede comprender una fracción funcional. Por ejemplo, la secuencia enlazadora puede comprender un aptámero, una ribozima, una horquilla que interactúa con proteínas, una matriz CRISPR, un intrón y un exón. La secuencia enlazadora puede comprender al menos aproximadamente 1, 2, 3, 4 o 5, o más fracciones funcionales. La secuencia enlazadora puede comprender, como máximo, aproximadamente 1, 2, 3, 4 o 5, o más fracciones funcionales.

En algunas realizaciones, el enlazador guía simple puede conectar el extremo 3' de la repetición CRISPR mínima al extremo 5' de la secuencia mínima de ARNcrtra. Como alternativa, el enlazador guía sencillo puede conectar el extremo 3' de la secuencia de ARNcrtra con el extremo 5' de la repetición CRISPR mínima. Es decir, un ácido nucleico guía sencillo puede comprender un segmento de unión a ADN de 5' unido a un segmento de unión a proteínas de 3'. Un ácido nucleico guía sencillo puede comprender un segmento de unión a proteínas de 5' unido a un segmento de unión a ADN de 3'.

Un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

Un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o

un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

Ácido nucleico guía doble

Un ácido nucleico dirigido a ácido nucleico puede ser un ácido nucleico guía doble. El ácido nucleico guía doble puede ser ARN. El ácido nucleico guía doble puede comprender dos moléculas de ácido nucleico separadas (es decir, polinucleótidos). Cada una de las dos moléculas de ácido nucleico de un ácido nucleico dirigido a ácido nucleico guía doble puede comprender un tramo de nucleótidos que pueden hibridarse entre sí de manera que los nucleótidos complementarios de las dos moléculas de ácido nucleico se hibridan para formar el dúplex bicatenario del segmento de unión a proteínas. Si no se especifica lo contrario, la expresión "ácido nucleico dirigido a ácido nucleico" puede ser inclusiva, refiriéndose tanto a los ácidos nucleicos dirigidos a ácido nucleico de una sola molécula como a los ácidos nucleicos dirigidos a ácido nucleico de doble molécula.

Un ácido nucleico dirigido a ácido nucleico guía doble puede comprender 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un ácido nucleico dirigido a ácido nucleico guía doble puede comprender 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

Complejo de un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio

Un ácido nucleico dirigido a ácido nucleico puede interactuar con un polipéptido dirigido a un sitio (por ejemplo, una nucleasa guiada por ácido nucleico, Cas9), formando así un complejo. El ácido nucleico dirigido a ácido nucleico puede guiar al polipéptido dirigido a un sitio a un ácido nucleico diana.

En algunas realizaciones, un ácido nucleico dirigido a ácido nucleico puede modificarse de forma tal que el complejo (por ejemplo, que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico) pueda unirse fuera del sitio de escisión del polipéptido dirigido a un sitio. En este caso, el ácido nucleico diana puede no interactuar con el complejo, y el ácido nucleico diana se puede escindir (por ejemplo, libre del complejo).

En algunas realizaciones, un ácido nucleico dirigido a ácido nucleico puede diseñado de modo que el complejo pueda unirse dentro del sitio de escisión del polipéptido dirigido a un sitio. En este caso, el ácido nucleico diana puede interactuar con el complejo y el ácido nucleico diana puede unirse (por ejemplo, unirse al complejo).

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprenda al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes* y, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una

secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y el ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de

identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y ácido nucleico dirigido a ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y el ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*)

o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y ácido nucleico dirigido a ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y el ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud

de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y un ácido nucleico dirigido a ácido nucleico guía doble que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y ácido nucleico dirigido a ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un complejo puede comprender un polipéptido dirigido a un sitio, en donde el polipéptido dirigido a

un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y el ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

Cualquier ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación puede ser recombinante, purificado y/o aislado.

Ácidos nucleicos que codifican un ácido nucleico dirigido a ácido nucleico y/o un polipéptido dirigido a un sitio

La presente divulgación proporciona un ácido nucleico que comprende una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación. En algunas realizaciones, el ácido nucleico que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación puede ser un vector (por ejemplo, un vector de expresión recombinante).

En algunas realizaciones, el vector de expresión recombinante puede ser una construcción vírica, (por ejemplo, una construcción de virus adenoasociado recombinante), una construcción adenovírica recombinante, una construcción lentivírica recombinante, una construcción retrovírica recombinante, etc.

Los vectores de expresión adecuados pueden incluir, pero sin limitación, vectores víricos (por ejemplo, vectores víricos basados en virus vaccinia, poliovirus, adenovirus, virus adenoasociados, SV40, virus del herpes simple, virus de la inmunodeficiencia humana, un vector retrovírico (por ejemplo, virus de la leucemia murina, virus de la necrosis del bazo y vectores procedentes de retrovirus, tales como el virus del sarcoma de Rous, virus del sarcoma de Harvey, virus de la leucosis aviar, un lentivirus, virus de la inmunodeficiencia humana, virus del sarcoma mieloproliferativo y virus del tumor mamario), vectores de plantas (por ejemplo, vector de ADN-T) y similares. Los siguientes vectores se pueden proporcionar, a modo de ejemplo, para las células hospedadoras eucariotas: pXT1, pSG5, pSVK3, pBPV, pMSG y pSVLSV40 (Pharmacia). Se pueden usar otros vectores siempre que sean compatibles con la célula hospedadora.

En algunos casos, un vector puede ser un vector linealizado. Un vector linealizado puede comprender un polipéptido dirigido a un sitio y/o un ácido nucleico dirigido a ácido nucleico. Un vector linealizado puede no ser un plásmido circular. Un vector linealizado puede comprender una rotura bicatenaria. Un vector linealizado puede comprender una secuencia que codifica una proteína fluorescente (por ejemplo, proteína fluorescente naranja (OFP)). Un vector linealizado puede comprender una secuencia que codifica un antígeno (por ejemplo, CD4). Un vector linealizado se puede linealizar (por ejemplo, cortar) en una región del vector que codifica partes del ácido nucleico dirigido a ácido nucleico. Por ejemplo, un vector linealizado se puede linealizar (por ejemplo, cortar) en una región del ácido nucleico dirigido a ácido nucleico a 5' a la parte de ARNcr del ácido nucleico dirigido a ácido nucleico. Un vector linealizado se puede linealizar (por ejemplo, cortar) en una región del ácido nucleico dirigido a ácido nucleico a 3' de la secuencia de

extensión del espaciador del ácido nucleico dirigido a ácido nucleico. Un vector linealizado se puede linealizar (por ejemplo, cortar) en una región del ácido nucleico dirigido a ácido nucleico que codifica la secuencia de ARNcr del ácido nucleico dirigido a ácido nucleico. En algunos casos, un vector linealizado o un vector superenrollado cerrado comprende una secuencia que codifica un polipéptido dirigido a un sitio (por ejemplo, Cas9), un promotor que impulsa la expresión de la secuencia que codifica el polipéptido dirigido a un sitio (por ejemplo, promotor CMV), una secuencia que codifica un enlazador (por ejemplo, 2A), una secuencia que codifica un marcador (por ejemplo, CD4 u OFP), una secuencia que codifica la parte de un ácido nucleico dirigido a ácido nucleico, un promotor que impulsa la expresión de la secuencia que codifica una parte del ácido nucleico dirigido a ácido nucleico, y una secuencia que codifica un marcador seleccionable (por ejemplo, ampicilina), o cualquier combinación de las mismas.

Un vector puede comprender un elemento de control de la transcripción y/o de la traducción. Dependiendo del hospedador/sistema de vector utilizado, se puede usar cualquiera de una serie de elementos de control de la transcripción y de la traducción adecuados, Incluyendo promotores constitutivos e inducibles, elementos potenciadores de la transcripción, terminadores de la transcripción, etc. en el vector de expresión.

En algunas realizaciones, una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación se puede unir operativamente a un elemento de control (por ejemplo, un elemento de control de la transcripción), tal como un promotor. El elemento de control de la transcripción puede ser funcional en una célula eucariota, (por ejemplo, una célula de mamífero), una célula procariota (por ejemplo, una célula bacteriana o de arquea). En algunas realizaciones, una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación se puede unir operativamente a múltiples elementos de control. El unión operativa a múltiples elementos de control puede permitir la expresión de la secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación en células procariotas o eucariotas.

Los ejemplos no limitantes de promotores eucariotas adecuados (es decir, promotores funcionales en una célula eucariota) pueden incluir los procedentes de citomegalovirus (CMV) inmediato precoz, timidina quinasa del virus del herpes simple (VHS), SV40 precoz y tardío, repeticiones terminales largas (LTR) de retrovirus, promotor del factor de elongación humano-1 (EF1), una construcción híbrida que comprende el potenciador del citomegalovirus (CMV) fusionado con el promotor activo beta (CAG) de pollo, promotor del virus de células madre murinas (MSCV), promotor de locus de fosfoglicerato quinasa 1 (PGK) y metalotioneína I de ratón. El promotor puede ser un promotor fúngico. El promotor puede ser un promotor vegetal. Se puede encontrar una base de datos de promotores vegetales (por ejemplo, PlantProm). El vector de expresión también puede contener un sitio de unión a ribosomas para el inicio de la traducción y un terminador de la transcripción. El vector de expresión también puede incluir secuencias apropiadas para amplificar la expresión. El vector de expresión también puede incluir secuencias de nucleótidos que codifican marcadores no nativos (por ejemplo, marcador 6xHis, marcador de hemaglutinina, proteína verde fluorescente, etc.) que se fusionan con el polipéptido dirigido a un sitio, dando lugar así a una proteína de fusión.

En algunas realizaciones, una secuencia o secuencias de nucleótidos que codifican un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación se puede unir operativamente a un promotor inducible (por ejemplo, promotor de choque térmico, promotor regulado por tetraciclina, promotor regulado por esteroides, promotor regulado por metales, promotor regulado por el receptor de estrógenos, etc.). En algunas realizaciones, una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesiten para llevar a cabo las realizaciones de los métodos de la divulgación un polipéptido de la divulgación se puede unir operativamente a un promotor constitutivo (por ejemplo, promotor CMV, promotor UBC). En algunas realizaciones, la secuencia de nucleótidos puede estar unida operativamente a un promotor restringido espacialmente y/o temporalmente (por ejemplo, un promotor específico del tejido, un promotor específico del tipo de célula, etc.).

Una secuencia o secuencias de nucleótidos que codifican un ácido nucleico dirigido a ácido nucleico de la divulgación, un polipéptido dirigido a un sitio de la divulgación, una proteína efectora, un agente de dirección genético multiplexado,

un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación se puede empaquetar dentro o sobre la superficie de los compartimentos biológicos para la administración a las células. Los compartimentos biológicos pueden incluir, pero sin limitación, virus (lentivirus, adenovirus), nanoesferas, liposomas, puntos cuánticos, nanopartículas, partículas de polietilenglicol, hidrogeles y micelas.

La introducción de los complejos, los polipéptidos y los ácidos nucleicos de la divulgación en las células pueden ocurrir por infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares.

Células y organismos transgénicos

La divulgación proporciona células y organismos transgénicos. El ácido nucleico de una célula hospedadora modificada genéticamente y/u organismo transgénico puede ser objeto de ingeniería genómica.

Las células ilustrativas que pueden usarse para generar células transgénicas de acuerdo con los métodos de la divulgación pueden incluir, pero sin limitación, célula HeLa, célula de ovario de hámster chino, célula 293-T, un feocromocitoma, un fibroblasto de neuroblastomas, un rabdomiosarcoma, una célula del ganglio de la raíz dorsal, una célula NSO, Tobacco BY-2, CV-I (ATCC CCL 70), COS-I (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C 1271 (ATCC CRL 1616), BS-C-I (ATCC CCL 26), MRC-5 (ATCC CCL 171), células L, HEK-293 (ATCC CRL1 573) y PC 12 (ATCC CRL-1721), HEK293T (ATCC CRL-11268), RBL (ATCC CRL-1378), SH-SY5Y (ATCC CRL-2266), MDCK (ATCC CCL-34), SJ-RH30 (ATCC CRL-2061), HepG2 (ATCC HB-8065), ND7/23 (ECACC 92090903), CHO (ECACC 85050302), Vera (ATCC CCL 81), Caco-2 (ATCC HTB 37), K562 (ATCC CCL 243), Jurkat (ATCC TIB-152), Per.Có, Huvec (ATCC PCS 100-010 primaria humana, CRL 2514 de ratón, CRL 2515, CRL 2516), HuH-7D12 (ECACC 01042712), 293 (ATCC CRL 10852), A549 (ATCC CCL 185), IMR-90 (ATCC CCL 186), MCF-7 (ATC HTB-22), U-2 OS (ATCC HTB-96) y T84 (ATCC CCL 248), o cualquier célula disponible en la colección Americana de Cultivos Tipo (ATCC), o cualquier combinación de las mismas.

Los organismos que pueden ser transgénicos pueden incluir bacterias, arqueas, eucariotas unicelulares, plantas, algas, hongos (por ejemplo, levadura), invertebrados (por ejemplo, mosca de la fruta, cnidario, equinodermo, nematodo, etc.), Vertebrados (por ejemplo, pez, anfibio, reptil, ave, mamífero), mamíferos (por ejemplo, un cerdo, una vaca, una cabra, una oveja, un roedor, una rata, un ratón, un primate no humano, un ser humano, etc.), etc.

Los organismos transgénicos pueden comprender células modificadas genéticamente. Los organismos transgénicos y/o las células modificadas genéticamente pueden comprender organismos y/o células que se han modificado genéticamente con un ácido nucleico exógeno que comprende una secuencia de nucleótidos que codifica el ácido nucleico dirigido a ácido nucleico de la divulgación, una proteína efectora y/o un polipéptido dirigido a un sitio, o cualquier combinación de los mismos.

Una célula modificada genéticamente puede comprender un polipéptido dirigido a un sitio exógeno y/o un ácido nucleico exógeno que comprende una secuencia de nucleótidos que codifica un polipéptido dirigido a un sitio. La expresión del polipéptido dirigido a un sitio en la célula puede llevar 0,1, 0,2, 0,5, 1, 2, 3, 4, 5, 6 o más días. Las células, introducidas con el polipéptido dirigido a un sitio, se pueden cultivar durante 0,1, 0,2, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o incluso más días antes de que las células puedan retirarse del cultivo celular y/o del organismo hospedador.

Sujetos

La divulgación proporciona la realización de los métodos de divulgación en un sujeto. Un sujeto puede ser un ser humano. Un sujeto puede ser un mamífero (por ejemplo, rata, ratón, vaca, perro, cerdo, oveja, caballo). Un sujeto puede ser un vertebrado o un invertebrado. Un sujeto puede ser un animal de laboratorio. Un sujeto puede ser un paciente. Un sujeto puede estar padeciendo una enfermedad. Un sujeto puede mostrar síntomas de una enfermedad. Un sujeto puede no mostrar síntomas de una enfermedad, pero,aún así, tener una enfermedad. Un sujeto puede estar bajo la atención médica de un cuidador (por ejemplo, el sujeto está hospitalizado y es tratado por un médico). Un sujeto puede ser una planta o un cultivo.

Kits

La presente divulgación proporciona kits para llevar a cabo los métodos de la divulgación. Un kit puede incluir uno o más de: un ácido nucleico dirigido a ácido nucleico de la divulgación, un polinucleótido que codifica un ácido nucleico dirigido a ácido nucleico, un polipéptido dirigido a un sitio de la divulgación, un polinucleótido que codifica un polipéptido dirigido a un sitio, una proteína efectora, un polinucleótido que codifica una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido que codifica un agente de dirección genético multiplexado, un polinucleótido

donante, una proteína de fusión en tándem, un polinucleótido que codifica una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación, o cualquier combinación de los mismos.

Un ácido nucleico dirigido a ácido nucleico de la divulgación, un polinucleótido que codifica un ácido nucleico dirigido a ácido nucleico, un polipéptido dirigido a un sitio de la divulgación, un polinucleótido que codifica un polipéptido dirigido a un sitio, una proteína efectora, un polinucleótido que codifica una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido que codifica un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un polinucleótido que codifica una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación se han descrito en detalle anteriormente.

Un kit puede comprender: (1) un vector que comprende una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico; y (2) un vector que comprende una secuencia de nucleótidos que codifica el polipéptido dirigido a un sitio y (2) un reactivo para la reconstitución y/o dilución de los vectores.

Un kit puede comprender: (1) un vector que comprende (i) una secuencia de nucleótidos que codifica un ácido nucleico dirigido un ácido nucleico, e (ii) una secuencia de nucleótidos que codifica el polipéptido dirigido a un sitio; y (2) un reactivo para la reconstitución y/o dilución del vector.

Un kit puede comprender: (1) un vector que comprende una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico; (2) un vector que comprende una secuencia de nucleótidos que codifica el polipéptido dirigido a un sitio; (3) un vector que comprende una secuencia de nucleótidos que codifica una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación; y (4) un reactivo para la reconstitución y/o dilución de los vectores.

Un kit puede comprender: (1) un vector que comprende (i) una secuencia de nucleótidos que codifica un ácido nucleico dirigido a ácido nucleico, (ii) una secuencia de nucleótidos que codifica un polipéptido dirigido a un sitio; (2) un vector que comprende una secuencia de nucleótidos que codifica una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación; y (3) un reactivo para la reconstitución y/o dilución de los vectores de expresión recombinantes.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra

que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes* y, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio puede comprender una secuencia de aminoácidos que comprenda al menos un 15 % de identidad de aminoácidos con un Cas9 de S. pyogenes, y dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC); y, un ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-

30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y un ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un enlazador que une al polipéptido dirigido a un sitio con una secuencia no nativa; y ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena

abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprenda al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido

nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprenda al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al

protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC), en donde el polipéptido dirigido a un sitio comprende una mutación en uno o ambos dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios nucleasa en al menos un 50 %; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; y una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprenda al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y un ácido nucleico dirigido a ácido nucleico guía doble (y/o un ácido nucleico que lo codifica) que comprende: 1) una primera molécula de ácido nucleico que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNc de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos y al menos 3 nucleótidos no emparejados de una protuberancia; y 2) una segunda molécula de ácido nucleico del ácido nucleico dirigido a ácido nucleico guía doble puede comprender una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos y al menos 1 nucleótido no emparejado de una protuberancia, en donde el 1 nucleótido no emparejado de la protuberancia está ubicado en la misma protuberancia que los 3 nucleótidos no emparejados de la repetición CRISPR mínima; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que codifica el mismo) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6, 7 u 8 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6, 7 u 8 nucleótidos contiguos y en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos, y comprende una región duplexada; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos. Este ácido nucleico dirigido a ácido nucleico puede denominarse ácido nucleico dirigido a ácido nucleico guía sencillo.

En algunos casos, un kit puede comprender un polipéptido dirigido a un sitio (y/o un ácido nucleico que lo codifica), en donde el polipéptido dirigido a un sitio comprende al menos un 15 % de identidad de aminoácidos con una Cas9 de *S. pyogenes* y dos dominios de escisión de ácido nucleico, en donde uno o ambos dominios de escisión de ácido nucleico comprenden al menos un 50 % de identidad de aminoácidos con un dominio nucleasa de Cas9 de *S. pyogenes*; y un ácido nucleico dirigido a ácido nucleico (y/o un ácido nucleico que lo codifica) que comprende una secuencia de extensión del espaciador de 10-5000 nucleótidos de longitud; una secuencia de espaciador de 12-30 nucleótidos de longitud, en donde el espaciador es al menos un 50 % complementario a un ácido nucleico diana; un dúplex que comprende: 1) una repetición CRISPR mínima que comprende al menos un 60 % de identidad con un ARNcr de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos y en donde la repetición CRISPR mínima tiene una longitud de 5-30 nucleótidos; 2) una secuencia mínima de ARNcrtra que comprende al menos un 60 % de identidad con un ARNcrtra de una bacteria (por ejemplo, *S. pyogenes*) sobre 6 nucleótidos contiguos y en donde la secuencia mínima de ARNcrtra tiene una longitud de 5-30 nucleótidos; y 3) una protuberancia en donde la protuberancia comprende al menos 3 nucleótidos no emparejados en la cadena de repetición CRISPR mínima del dúplex y al menos 1 nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex; una secuencia enlazadora que une la repetición CRISPR mínima y el ARNcrtra mínimo y comprende una longitud de 3-5000 nucleótidos; un ARNcrtra 3' que comprende al menos un 60 % de identidad con un ARNcrtra de un procariota (por ejemplo, *S. pyogenes*) o un fago sobre 6 nucleótidos contiguos, en donde el ARNcrtra 3' comprende una longitud de 10-20 nucleótidos y comprende una región duplexada; un dominio P que comienza desde 1-5 nucleótidos cadena abajo del dúplex que comprende la repetición CRISPR mínima y el ARNcrtra mínimo, comprende 1-10 nucleótidos, comprende una secuencia que puede hibridarse con un motivo adyacente al protoespaciador en un ácido nucleico diana, puede formar una horquilla y se encuentra en la región de ARNcrtra 3'; y/o una extensión de ARNcrtra que comprende 10-5000 nucleótidos de longitud, o cualquier combinación de los mismos.

En algunas realizaciones de cualquiera de los kits anteriores, el kit puede comprender un ácido nucleico dirigido a ácido nucleico guía sencillo. En algún aspecto de cualquiera de los kits anteriores, el kit puede comprender un ácido nucleico dirigido a ácido nucleico guía doble. En algunas realizaciones de cualquiera de los kits anteriores, el kit puede comprender dos o más ácidos nucleicos dirigidos a ácido nucleicos guía sencillos. En algunas realizaciones, un vector puede codificar un ácido nucleico dirigido a ácido nucleico.

En algunas realizaciones de cualquiera de los kits anteriores, el kit puede comprender además un polinucleótido donante o una secuencia polinucleotídica que codifica el polinucleótido donante, para efectuar la modificación genética deseada. Los componentes de un kit pueden estar en recipientes separados; o se pueden combinar en un solo recipiente.

Un kit descrito anteriormente comprende además uno o más reactivos adicionales, donde dichos reactivos adicionales se pueden seleccionar entre: un tampón, un tampón para introducir un elemento polipeptídico o polinucleótido del kit en una célula, un tampón de lavado, un reactivo de control, un vector de control, un polinucleótido de ARN de control, un reactivo para la producción *in vitro* del polipéptido a partir de ADN, adaptadores para la secuenciación y similares. Un tampón puede ser un tampón de estabilización, un tampón reconstituyente o un tampón diluyente.

En algunos casos, un kit puede comprender uno o más reactivos adicionales específicos para plantas y/u hongos. Los uno o más reactivos adicionales para plantas y/u hongos pueden incluir, por ejemplo, suelo, nutrientes, plantas, semillas, esporas, agrobacterias, vector de ADN-T y un vector pBINAR.

Además de los componentes mencionados anteriormente, un kit puede incluir además instrucciones para usar los componentes del kit para poner en práctica los métodos. Las instrucciones para poner en práctica los métodos se registran generalmente en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden imprimirse sobre un sustrato, tal como papel o plástico, etc. Las instrucciones pueden estar presentes en los kits en forma de prospecto, en la etiqueta del envase del kit o los componentes del mismo (es decir, asociado al envase o envase secundario), etc. Las instrucciones pueden estar presentes en forma de un archivo de datos almacenado electrónicamente presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, unidad flash, etc. En algunos casos, las instrucciones reales no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota (por ejemplo, a través de Internet). Un ejemplo de esta realización es un kit que incluye una página web donde pueden verse las instrucciones y/o a través de la cual pueden descargarse las instrucciones. Al igual que con las instrucciones, esto significa que, para obtenerse las instrucciones, puede registrarse en un sustrato adecuado.

En algunas realizaciones, un kit puede comprender un vector linealizado. Un vector linealizado puede comprender un plásmido que comprenda un polipéptido dirigido a un sitio y/o un ácido nucleico dirigido a ácido nucleico que esté linealizado (por ejemplo, que no sea circular). Un vector linealizado puede almacenarse en un tampón que comprenda Tris-HCl 10 mM, pH 8,0 y EDTA 1 mM, pH 8,0. Un kit puede comprender aproximadamente 20 microlitros del vector de nucleasa CRISPR linealizado. En algunas realizaciones, un kit puede comprender uno o más vectores circulares.

En algunas realizaciones, un kit puede comprender un tampón de hibridación de oligonucleótidos. Un tampón de hibridación de oligonucleótidos puede ser un tampón usado para hibridar oligos de ADN entre sí para generar un ADN bicatenario que codifique un ácido nucleico dirigido a ácido nucleico. Un tampón de hibridación de oligonucleótidos puede estar al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más concentrado que la concentración de uso. Un tampón de hibridación de oligonucleótidos puede estar 10 veces más concentrado que la concentración cuando se usa. Un tampón de hibridación de oligonucleótidos puede comprender Tris-HCl 100 mM, pH 8,0, EDTA 10 mM, pH 8,0 y NaCl 1 M. Un kit puede comprender 250 microlitros del tampón de hibridación de oligonucleótidos.

Un kit puede comprender agua exenta de DNasa. Un kit puede comprender agua exenta de ARNasa. Un kit puede comprender al menos 1,5 mililitros de agua exenta de RNasa y/o DNAsa.

Un kit puede comprender un tampón de ligadura. Se puede usar un tampón de ligadura para ligar oligonucleótidos al vector de nucleasa CRISPR linealizado. Un tampón de ligadura puede estar al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más concentrado que la concentración de uso. Un tampón de ligadura puede ser 5 veces más concentrado que la concentración de uso. Un tampón de ligadura x5 puede comprender Tris-HCl 250 mM, pH 7,6, $MgCl_2$ 50 mM, ATP 5 mM, DTT 5 mM, y polietilenglicol-8000 al 25 % (p/v). Un kit puede comprender aproximadamente 80 microlitros de un tampón de ligadura.

Un kit puede comprender una ADN ligasa. Se puede usar una ADN ligasa para ligar los oligonucleótidos al vector de nucleasa CRISPR linealizado. Una ADN ligasa puede comprender Tris-HCl 10 mM, pH 7,5, KCl 50 mM, DTT 1 mM y glicerol al 50 % (v/v). Un kit puede comprender 20 microlitros de una ADN ligasa.

Un kit puede comprender un cebador de secuenciación. El cebador de secuenciación puede usarse para secuenciar el vector una vez que los oligonucleótidos se han ligado en un vector linealizado. Un cebador de secuenciación se puede diluir en tampón Tris-EDTA a pH 8,0. Un kit puede comprender 20 microlitros de un cebador de secuenciación.

Un kit puede comprender un oligonucleótido de control. Un oligonucleótido de control puede ser un oligonucleótido que se vaya a ligar en un vector linealizado, pero que no codifica un ácido nucleico dirigido a ácido nucleico. Un oligonucleótido de control puede diluirse a una concentración x1 del tampón de hibridación de oligonucleótidos. Un kit puede comprender 10 microlitros de un oligonucleótido de control.

En algunos casos, un kit puede comprender un vector linealizado que comprenda un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, un tampón de hibridación de oligonucleótidos, agua exenta de DNAsa/RNAsa, un tampón de ligadura, una enzima ligasa, un cebador de secuenciación y un oligonucleótido de control, o cualquier combinación de los mismos.

Composiciones farmacéuticas

Las moléculas, tales como un ácido nucleico dirigido a ácido nucleico de la divulgación como el descrito en el presente documento, un polinucleótido que codifica un ácido nucleico dirigido a ácido nucleico, un polipéptido dirigido a un sitio de la divulgación, un polinucleótido que codifica un polipéptido dirigido a un sitio, una proteína efectora, un polinucleótido que codifica una proteína efectora, un agente de dirección genético multiplexado, un polinucleótido que codifica un agente de dirección genético multiplexado, un polinucleótido donante, una proteína de fusión en tándem, un polinucleótido que codifica una proteína de fusión en tándem, un elemento indicador, un elemento genético de interés, un componente de un sistema dividido y/o cualquier ácido nucleico o molécula proteica que se necesite para llevar a cabo las realizaciones de los métodos de la divulgación, se pueden formular en una composición farmacéutica.

Una composición farmacéutica puede comprender una combinación de cualquier molécula descrita en el presente documento con otros componentes químicos, tales como vehículos, estabilizantes, diluyentes, agentes dispersantes, agentes de suspensión, agentes espesantes y/o excipientes. La composición farmacéutica puede facilitar la administración de la molécula a un organismo. Las composiciones farmacéuticas pueden administrarse en cantidades terapéuticamente eficaces como composiciones farmacéuticas mediante diversas formas y vías que incluyen, por ejemplo, administración intravenosa, subcutánea, intramuscular, oral, rectal, aerosol, parenteral, oftálmica, pulmonar, transdérmica, vaginal, ótica, nasal y tópica.

Una composición farmacéutica puede administrarse de manera local o sistémica, por ejemplo, mediante la inyección de la molécula directamente en un órgano, opcionalmente, en una formulación en depósito o de liberación sostenida. Las composiciones farmacéuticas pueden proporcionarse en forma de una formulación de liberación rápida, en forma de una formulación de liberación prolongada o en forma de una formulación de liberación intermedia. Una forma de liberación rápida puede proporcionar una liberación inmediata. Una formulación de liberación prolongada puede proporcionar una liberación controlada o una liberación retardada sostenida.

Para la administración oral, las composiciones farmacéuticas pueden formularse fácilmente combinando las moléculas con vehículos o excipientes farmacéuticamente aceptables. Dichos vehículos pueden usarse para formular comprimidos, polvos, píldoras, grageas, cápsulas, líquidos, geles, jarabes, elixires, pastas, suspensiones y similares, para la ingestión oral por parte de un sujeto.

Se pueden obtener preparados farmacéuticos para uso oral mezclando uno o más excipientes sólidos con una o más de las moléculas descritas en el presente documento, moliendo opcionalmente la mezcla resultante y procesando la mezcla de gránulos, después de añadir adyuvantes adecuados, si se desea, para obtenerse comprimidos o núcleos de grageas. Los núcleos se pueden proporcionar con recubrimientos adecuados. Para este fin, pueden usarse soluciones de azúcar concentradas, que pueden contener un excipiente tal como goma arábiga, talco, polivinilpirrolidona, gel de carbopol, polietilenglicol y/o dióxido de titanio, soluciones de laca y disolventes orgánicos adecuados o mezclas de disolventes. Se pueden añadir colorantes o pigmentos a los comprimidos o recubrimientos de grageas, por ejemplo, para identificar o para caracterizar diferentes combinaciones de dosis de compuestos activos.

Los preparados farmacéuticos que se pueden usar por vía oral incluyen cápsulas de ajuste por presión hechas de gelatina, así como cápsulas blandas, selladas fabricadas de gelatina y un plastificante, tal como glicerol o sorbitol. En algunos aspectos, la cápsula comprende una cápsula de gelatina dura que comprende uno o más productos farmacéuticos, gelatinas bovinas y vegetales. Una gelatina puede ser procesada en condiciones alcalinas. Las cápsulas de ajuste a presión pueden comprender los principios activos mezclados con carga, tal como lactosa, aglutinantes, tales como almidones y/o lubricantes, tales como talco o estearato de magnesio y, estabilizantes. En las cápsulas blandas, la molécula puede disolverse o suspenderse en líquidos adecuados, tales como aceites grasos, parafina líquida o polietilenglicoles líquidos. Pueden añadirse estabilizantes. Todas las formulaciones para administración oral se proporcionan en dosis adecuadas para dicha administración.

Para la administración bucal o sublingual, las composiciones pueden ser comprimidos, pastillas para chupar o geles.

Las inyecciones parentales pueden formularse para inyección en bolo o infusión continua. Las composiciones farmacéuticas pueden estar en una forma adecuada para inyección parenteral tal como una suspensión, solución o emulsión estéril en vehículos aceitosos o acuosos, y pueden contener agentes de formulación tales como agentes de suspensión, estabilizantes y/o dispersantes. Las formulaciones farmacéuticas para su administración parenteral

pueden incluir soluciones acuosas de los compuestos activos en forma hidrosoluble.

Se pueden preparar suspensiones de las moléculas en forma de suspensiones oleosos para inyección. Los disolventes o vehículos lipófilos adecuados incluyen aceites grasos, tales como aceite de sésamo, o ésteres de ácidos grasos sintéticos, tales como oleato de etilo o triglicéridos, o liposomas. Las suspensiones de inyección acuosa pueden contener sustancias que aumentan la viscosidad de la emulsión, tales como carboximetilcelulosa sódica, sorbitol o dextrano. La suspensión también puede contener estabilizantes o agentes adecuados que aumenten la solubilidad de las moléculas para permitir la preparación de soluciones muy concentradas. Como alternativa, el principio activo se puede encontrar en forma de polvo para su constitución con un vehículo adecuado, por ejemplo, agua apirógena estéril, antes de su uso.

Los compuestos activos pueden administrarse por vía tópica y pueden formularse en varias composiciones que se pueden administrar por vía tópica, tales como soluciones, suspensiones, lociones, geles, pastas, palos medicinales, bálsamos, cremas y pomadas. Dichas composiciones farmacéuticas pueden comprender solubilizantes, estabilizantes, agentes potenciadores de la tonicidad, tampones y conservantes.

Las formulaciones adecuadas para la administración transdérmica de las moléculas pueden emplear dispositivos de administración transdérmica y parches de administración transdérmica, y pueden ser emulsiones lipófilas o soluciones acuosas tamponadas, disueltas y/o dispersadas en un polímero o un adhesivo. Dichos parches pueden construirse para la administración continua, pulsátil o a demanda de moléculas. La administración transdérmica se puede lograr por medio de parches iontoforéticos y similares. Además, los parches transdérmicos pueden proporcionar una administración controlada. Se puede reducir la velocidad de absorción usando membranas de control de la velocidad o atrapando el compuesto dentro de una matriz de polímero o gel. Por el contrario, se pueden usar potenciadores de la absorción para aumentar la absorción. Un potenciador o vehículo de absorción puede incluir disolventes absorbibles farmacéuticamente aceptables para ayudar al paso a través de la piel. Por ejemplo, los dispositivos transdérmicos pueden tener la forma de un vendaje que comprenda un miembro de soporte, un depósito que contenga compuestos y vehículos, una barrera de control de la velocidad para administrar los compuestos en la piel del sujeto a una velocidad controlada y predeterminada durante un período prolongado de tiempo, y adhesivos para fijar el dispositivo a la piel.

Para la administración por inhalación, la molécula puede estar en forma de aerosol, una nebulización o un polvo. Las composiciones farmacéuticas pueden administrarse en forma de presentación en aerosol a partir de envases presurizados o un nebulizador, con el uso de un propulsor adecuado, por ejemplo, diclorodifluorometano, triclorofluorometano, diclorotetrafluoroetano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad medida. Las cápsulas y los cartuchos de, por ejemplo, gelatina para su uso en un inhalador o insuflador pueden formularse para que contengan una mezcla de polvo de los compuestos y una base de polvo adecuada, tal como lactosa o almidón.

Las moléculas también se pueden formular en composiciones rectales tales como enemas, geles rectales, espumas rectales, aerosoles rectales, supositorios, supositorios de gelatina o enemas de retención, que contiene bases para supositorio convencionales, tales como manteca de cacao u otros glicéridos, así como polímeros sintéticos tales como polivinilpirrolidona y PEG. En las formas en supositorio de las composiciones, se puede usar una cera de bajo punto de fusión, tal como una mezcla de glicéridos de ácidos grasos o manteca de cacao.

Al poner en práctica los métodos de la divulgación, se pueden administrar cantidades terapéuticamente eficaces de los compuestos descritos en el presente documento en composiciones farmacéuticas a un sujeto que tenga una enfermedad o afección que haya que tratar. Una cantidad terapéuticamente eficaz puede variar ampliamente dependiendo de la gravedad de la enfermedad, la edad y salud relativa del sujeto, la potencia de los compuestos usados, así como de otros factores. Los compuestos se pueden usar solos o en combinación con uno o más agentes terapéuticos como componentes de mezclas.

Las composiciones farmacéuticas pueden formularse usando uno o más vehículos fisiológicamente aceptables que comprenden excipientes y adyuvantes, que faciliten el procesamiento de la molécula en preparados que pueden usarse farmacéuticamente. La formulación puede modificarse dependiendo de la vía de administración escogida. Las composiciones farmacéuticas que comprenden una molécula descrita en el presente documento pueden fabricarse, por ejemplo, mediante procesos de mezcla, disolución, granulación, preparación de grageas, levigación, emulsionado, encapsulación, atrapamiento o compresión.

Las composiciones farmacéuticas pueden incluir al menos un vehículo, diluyente o excipiente farmacéuticamente aceptable, y la molécula descrita en el presente documento en forma de base libre o de sal farmacéuticamente aceptable. Los métodos y las composiciones farmacéuticas descritos en el presente documento incluyen el uso de formas cristalinas (también conocidas como polimorfos) y metabolitos activos de estos compuestos que tienen el mismo tipo de actividad.

Los métodos para la preparación de composiciones que comprenden los compuestos descritos en el presente documento pueden incluir la formulación de la molécula con uno o más excipientes o vehículos farmacéuticamente

aceptables, inertes, para formar una composición sólida, semisólida o líquida. Las composiciones sólidas pueden incluir, por ejemplo, polvos, comprimidos, gránulos dispersables, cápsulas, obleas y supositorios. Las composiciones líquidas pueden incluir, por ejemplo, soluciones en las que hay disuelto un compuesto, emulsiones que comprenden un compuesto o una solución que contiene liposomas, micelas o nanopartículas que comprenden un compuesto tal como se ha desvelado en el presente documento. Las composiciones semisólidas pueden incluir, por ejemplo, geles, suspensiones y cremas. Las composiciones pueden estar en soluciones o suspensiones líquidas, en formas sólidas adecuadas para su disolución o suspensión en un líquido antes de su uso o como emulsiones. Estas composiciones también pueden contener cantidades menores de sustancias auxiliares, no tóxicas, tales como agentes humectantes o emulsionantes, agentes tamponadores de pH y otros aditivos farmacéuticamente aceptables.

Los ejemplos no limitantes de formas farmacéuticas pueden incluir piensos, alimentos, microgránulo, pastilla para chupar, líquido, elixir, aerosol, inhalador, pulverización, polvo, comprimido, píldora, cápsula, gel, comprimidos de gel, nanosuspensión, nanopartícula, microgel, trociscos en supositorio, suspensiones acuosas u oleosas, pomada, parche, loción, dentífrico, emulsión, cremas, gotas, polvos o gránulos dispersables, emulsión en cápsulas de gel duras o blandas, jarabes, fitofármacos y nutracéuticos, o cualquier combinación de los mismos.

Los ejemplos no limitantes de excipientes farmacéuticamente aceptables pueden incluir agentes de granulación, agentes aglutinantes, agentes lubricantes, agentes disgregantes, agentes edulcorantes, sustancias de deslizamiento, antiadherentes, agentes antiestáticos, tensioactivos, antioxidantes, gomas, agentes de recubrimiento, agentes colorantes, agentes aromatizantes, agentes de recubrimiento, plastificantes, conservantes, agentes de suspensión, agentes emulsionantes, material celulósico vegetal y agentes de esferonización, o cualquier combinación de los mismos.

Una composición puede ser, por ejemplo, una forma de liberación inmediata o una formulación de liberación controlada. Se puede formular una formulación de liberación inmediata para permitir que las moléculas actúen rápidamente. Los ejemplos no limitantes de formulaciones de liberación inmediata pueden incluir formulaciones fácilmente solubles. Una formulación de liberación controlada puede ser una formulación farmacéutica que se haya adaptado para que las velocidades de liberación del fármaco y los perfiles de liberación del fármaco sean acordes a las necesidades fisiológicas y cronoterapéuticas o, como alternativa, que se haya formulado para efectuar la liberación de un fármaco a una velocidad programada. Los ejemplos no limitantes de formulaciones de liberación controlada pueden incluir gránulos, gránulos de liberación retardada, hidrogeles (por ejemplo, de origen sintético o natural), otros agentes gelificantes (por ejemplo, fibras dietéticas formadoras de gel), formulaciones basadas en matriz (por ejemplo, formulaciones que comprenden un material polimérico que tiene al menos un principio activo disperso), gránulos dentro de una matriz, mezclas poliméricas, masas granulares y similares.

Una formulación de liberación controlada puede ser una forma de liberación retardada. Se puede formular una forma de liberación retardada para retrasar la acción de una molécula durante un período prolongado de tiempo. Se puede formular una forma de liberación retardada para retrasar la liberación de una dosis eficaz de una o más moléculas, por ejemplo, durante aproximadamente 4, aproximadamente 8, aproximadamente 12, aproximadamente 16 o aproximadamente 24 horas.

Una formulación de liberación controlada puede ser una forma de liberación sostenida. Se puede formular una forma de liberación sostenida para mantener, por ejemplo, la acción de la molécula durante un período prolongado de tiempo. Se puede formular una forma de liberación sostenida para proporcionar una dosis eficaz de cualquier molécula descrita en el presente documento (por ejemplo, proporcionar un perfil sanguíneo fisiológicamente eficaz) durante aproximadamente 4, aproximadamente 8, aproximadamente 12, aproximadamente 16 o aproximadamente 24 horas.

Métodos de administración y métodos de tratamiento.

Las composiciones farmacéuticas que contienen moléculas descritas en el presente documento pueden administrarse para tratamientos profilácticos y/o terapéuticos. En las aplicaciones terapéuticas, las composiciones pueden administrarse a un sujeto que ya padece una enfermedad o afección, en una cantidad suficiente para curar o al menos detener parcialmente los síntomas de la enfermedad o afección, o para curar, sanar, o mejorar la afección. Las cantidades eficaces para este uso pueden variar según la gravedad y el curso de la enfermedad o afección, la terapia previa, el estado de salud del sujeto, el peso y la respuesta a los fármacos y el juicio del médico tratante.

Se pueden administrar múltiples agentes terapéuticos en cualquier orden o simultáneamente. Si se administran simultáneamente, los múltiples agentes terapéuticos se pueden proporcionar en una sola forma unificada o en múltiples formas, por ejemplo, como múltiples píldoras separadas. Las moléculas se pueden acondicionar juntas o por separado, en un solo envase o en una pluralidad de envases. Uno o todos los agentes terapéuticos pueden administrarse en múltiples dosis. Si no se administran simultáneamente, el tiempo entre las múltiples dosis puede variar hasta aproximadamente un mes.

Las moléculas descritas en el presente documento pueden administrarse antes, durante o después de la aparición de una enfermedad o afección, y el momento de la administración de la composición que contiene un compuesto puede variar. Por ejemplo, las composiciones farmacéuticas pueden usarse como composiciones profilácticas y pueden

administrarse de manera continua a sujetos con una propensión a afecciones o enfermedades a fin de prevenir la aparición de la enfermedad o afección. Las moléculas y las composiciones farmacéuticas se pueden administrar a un sujeto durante o tan pronto como sea posible tras el inicio de los síntomas. La administración de las moléculas puede iniciarse dentro de las primeras 48 horas del inicio de los síntomas, dentro de las primeras 24 horas del inicio de los síntomas, dentro de las primeras 6 horas del inicio de los síntomas o dentro de las 3 horas del inicio de los síntomas. La administración inicial puede ser por cualquier vía práctica, tal como por cualquier vía descrita en el presente documento usando cualquier formulación descrita en el presente documento. Se puede administrar una molécula tan pronto como sea posible tras la detección o sospecha del inicio de una enfermedad o afección, y durante el tiempo necesario para el tratamiento de la enfermedad, tal como, por ejemplo, de aproximadamente 1 mes a aproximadamente 3 meses. La duración del tratamiento puede variar para cada sujeto.

Una molécula se puede acondicionar en un compartimento biológico. Un compartimento biológico que comprende la molécula puede administrarse a un sujeto. Los compartimentos biológicos pueden incluir, pero sin limitación, virus (lentivirus, adenovirus), nanoesferas, liposomas, puntos cuánticos, nanopartículas, micropartículas, nanocápsulas, vesículas, partículas de polietilenglicol, hidrogeles y micelas.

Por ejemplo, un compartimento biológico puede comprender un liposoma. Un liposoma puede ser una estructura de autoensamblaje que comprenda una o más bicapas lipídicas, pudiendo comprender cada una de las cuales dos monocapas que contengan moléculas de lípidos anfipáticos con orientación opuesta. Los lípidos anfipáticos pueden comprender un grupo de cabeza polar (hidrófilo) unido covalentemente a una o dos o más cadenas de acilo o alquilo no polares (hidrófobas). Los contactos energéticamente desfavorables entre las cadenas de acilo hidrófobas y un medio acuoso circundante inducen a las moléculas de lípidos anfipáticos a organizarse de manera que los grupos de cabeza polares puedan orientarse hacia la superficie de la bicapa y las cadenas de acilo estén orientadas hacia el interior de la bicapa, protegiendo de manera eficaz las cadenas de acilo del contacto con el ambiente acuoso.

Los ejemplos de compuestos anfipáticos preferidos usados en liposomas pueden incluir fosfoglicéridos y esfingolípidos, cuyos ejemplos representativos incluyen fosfatidilcolina, fosfatidiletanolamina, fosfatidilserina, fosfatidilinositol, ácido fosfatídico, fosfatidilglicerol, palmitoiloleoil fosfatidilcolina, lisofosfatidilcolina, lisofosfatidiletanolamina, dimiristoilfosfatidilcolina (DMPC), dipalmitoilfosfatidilcolina (DPPC), dioleoilfosfatidilcolina, diestearoilfosfatidilcolina (DSPC), dilinoleoilfosfatidilcolina y esfingomielina de huevo, o cualquier combinación de los mismos.

Un compartimento biológico puede comprender una nanopartícula. Una nanopartícula puede comprender un diámetro de aproximadamente 40 nanómetros a aproximadamente 1,5 micrómetros, de aproximadamente 50 nanómetros a aproximadamente 1,2 micrómetros, de aproximadamente 60 nanómetros a aproximadamente 1 micrómetro, de aproximadamente 70 nanómetros a aproximadamente 800 nanómetros, de aproximadamente 80 nanómetros a aproximadamente 600 nanómetros, de aproximadamente 90 nanómetros a aproximadamente 400 nanómetros, de aproximadamente 100 nanómetros a aproximadamente 200 nanómetros.

En algunos casos, a medida que aumenta el tamaño de la nanopartícula, puede reducirse o prolongarse la velocidad de liberación, y a medida que disminuye el tamaño de la nanopartícula, se puede aumentar la velocidad de liberación.

La cantidad de albúmina en las nanopartículas puede variar del aproximadamente 5 % al aproximadamente 85 % de albúmina (v/v), del aproximadamente 10 % al aproximadamente 80 %, del aproximadamente 15 % al aproximadamente 80 %, del aproximadamente 20 % al aproximadamente 70 % de albúmina (v/v), del aproximadamente 25 % al aproximadamente 60 %, del aproximadamente 30 % al aproximadamente 50 % o del aproximadamente 35 % al aproximadamente 40 %. La composición farmacéutica puede comprender hasta 30, 40, 50, 60, 70 u 80 % o más de nanopartícula. En algunos casos, las moléculas de ácido nucleico de la divulgación pueden unirse a la superficie de la nanopartícula.

Un compartimento biológico puede comprender un virus. El virus puede ser un sistema de administración para las composiciones farmacéuticas de la divulgación. Los virus ilustrativos pueden incluir lentivirus, retrovirus, adenovirus, virus del herpes simple I o II, parvovirus, virus de la reticuloendoteliosis y virus adenoasociados (AAV). Las composiciones farmacéuticas de la divulgación pueden administrarse a una célula usando un virus. El virus puede infectar y transducir la célula *in vivo*, *ex vivo* o *in vitro*. En la administración *ex vivo* e *in vitro*, las células transducidas pueden administrarse a un sujeto que necesite terapia.

Las composiciones farmacéuticas se pueden acondicionar en sistemas de administración vírica. Por ejemplo, las composiciones pueden acondicionarse en viriones mediante un sistema de acondicionamiento libre de virus auxiliar HSV-1.

Los sistemas de administración vírica (por ejemplo, los virus que comprenden las composiciones farmacéuticas de la divulgación) se pueden administrar mediante inyección directa, inyección estereotáxica, vía intracerebroventricular, mediante sistemas de infusión de minibomba, por convección, catéteres, administración intravenosa, parenteral, inyección intraperitoneal y/o subcutánea, a una célula, un tejido o un órgano de un sujeto necesitado. En algunos casos, las células pueden transducirse *in vitro* o *ex vivo* con sistemas de administración vírica. Las células transducidas

pueden administrarse a un sujeto que tenga una enfermedad. Por ejemplo, se puede transducir una célula madre con un sistema de administración vírica que comprenda una composición farmacéutica, y la célula madre puede implantarse en el paciente para tratar una enfermedad. En algunos casos, la dosis de células transducidas administrada a un sujeto puede ser de aproximadamente 1×10^5 células/kg, aproximadamente 5×10^5 células/kg, aproximadamente 1×10^6 células/kg, aproximadamente 2×10^6 células/kg, aproximadamente 3×10^6 células/kg, aproximadamente 4×10^6 células/kg, aproximadamente 5×10^6 células/kg, aproximadamente 6×10^6 células/kg, aproximadamente 7×10^6 células/kg, aproximadamente 8×10^6 células/kg, aproximadamente 9×10^6 células/kg, aproximadamente 1×10^7 células/kg, aproximadamente 5×10^7 células/kg, aproximadamente 1×10^8 células/kg o más en una sola dosis.

Las composiciones farmacéuticas en compartimentos biológicos pueden usarse para tratar enfermedades inflamatorias tales como artritis, cánceres, tales como, por ejemplo, cáncer de hueso, cáncer de mama, cáncer de piel, cáncer de próstata, cáncer de hígado, cáncer de pulmón, cáncer de garganta y cáncer de riñón, infecciones bacterianas, para tratar el daño nervioso, enfermedades pulmonares, hepáticas y renales, tratamiento de los ojos, lesiones de la médula espinal, cardiopatía, enfermedad arterial.

La introducción de los compartimentos biológicos en las células puede ocurrir por infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares.

Dosificación

Las composiciones farmacéuticas descritas en el presente documento pueden estar en formas de dosificación unitarias adecuadas para la administración única de dosis precisas. En la forma de dosificación unitaria, la formulación se puede dividir en dosis unitarias que contengan cantidades apropiadas de uno o más compuestos. La dosis unitaria puede estar en forma de un envase que contenga cantidades diferenciadas de la formulación. Los ejemplos no limitantes pueden incluir comprimidos o cápsulas acondicionados y polvos en viales o ampollas. Las composiciones de suspensión acuosa se pueden acondicionar en recipientes monodosis no reutilizables. Se pueden usar recipientes reutilizables multidosis, por ejemplo, en combinación con un conservante. Las formulaciones para inyección parenteral se pueden presentar en forma de dosificación unitaria, por ejemplo, en ampollas o en envases multidosis con un conservante.

Una molécula descrita en el presente documento puede estar presente en una composición en un intervalo de aproximadamente 1 mg a aproximadamente 2000 mg; de aproximadamente 5 mg a aproximadamente 1000 mg, de aproximadamente 10 mg a aproximadamente 25 mg a 500 mg, de aproximadamente 50 mg a aproximadamente 250 mg, de aproximadamente 100 mg a aproximadamente 200 mg, de aproximadamente 1 mg a aproximadamente 50 mg, de aproximadamente 50 mg a aproximadamente 100 mg, de aproximadamente 100 mg a aproximadamente 150 mg, de aproximadamente 150 mg a aproximadamente 200 mg, de aproximadamente 200 mg a aproximadamente 250 mg, de aproximadamente 250 mg a aproximadamente 300 mg, de aproximadamente 300 mg a aproximadamente 350 mg, de aproximadamente 350 mg a aproximadamente 400 mg, de aproximadamente 400 mg a aproximadamente 450 mg, de aproximadamente 450 mg a aproximadamente 500 mg, de aproximadamente 500 mg a aproximadamente 550 mg, de aproximadamente 550 mg a aproximadamente 600 mg, de aproximadamente 600 mg a aproximadamente 650 mg, de aproximadamente 650 mg a aproximadamente 700 mg, de aproximadamente 700 mg a aproximadamente 750 mg, de aproximadamente 750 mg a aproximadamente 800 mg, de aproximadamente 800 mg a aproximadamente 850 mg, de aproximadamente 850 mg a aproximadamente 900 mg, de aproximadamente 900 mg a aproximadamente 950 mg o de aproximadamente 950 mg a aproximadamente 1000 mg.

Una molécula descrita en el presente documento puede estar presente en una composición en una cantidad de aproximadamente 1 mg, aproximadamente 2 mg, aproximadamente 3 mg, aproximadamente 4 mg, aproximadamente 5 mg, aproximadamente 10 mg, aproximadamente 15 mg, aproximadamente 20 mg, aproximadamente 25 mg, aproximadamente 30 mg, aproximadamente 35 mg, aproximadamente 40 mg, aproximadamente 45 mg, aproximadamente 50 mg, aproximadamente 55 mg, aproximadamente 60 mg, aproximadamente 65 mg, aproximadamente 70 mg, aproximadamente 75 mg, aproximadamente 80 mg, aproximadamente 85 mg, aproximadamente 90 mg, aproximadamente 95 mg, aproximadamente 100 mg, aproximadamente 125 mg, aproximadamente 150 mg, aproximadamente 175 mg, aproximadamente 200 mg, aproximadamente 250 mg, aproximadamente 300 mg, aproximadamente 350 mg, aproximadamente 400 mg, aproximadamente 450 mg, aproximadamente 500 mg, aproximadamente 550 mg, aproximadamente 600 mg, aproximadamente 650 mg, aproximadamente 700 mg, aproximadamente 750 mg, aproximadamente 800 mg, aproximadamente 850 mg, aproximadamente 900 mg, aproximadamente 950 mg, aproximadamente 1000 mg, aproximadamente 1050 mg, aproximadamente 1100 mg, aproximadamente 1150 mg, aproximadamente 1200 mg, aproximadamente 1250 mg, aproximadamente 1300 mg, aproximadamente 1350 mg, aproximadamente 1400 mg, aproximadamente 1450 mg, aproximadamente 1500 mg, aproximadamente 1550 mg, aproximadamente 1600 mg, aproximadamente 1650 mg, aproximadamente 1700 mg, aproximadamente 1750 mg, aproximadamente 1800 mg, aproximadamente 1850 mg, aproximadamente 1900 mg, aproximadamente 1950 mg o aproximadamente 2000 mg.

Una molécula (por ejemplo, un polipéptido dirigido a un sitio, ácido nucleico dirigido a ácido nucleico y/o complejo de un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico) descrita en el presente documento puede estar presente en una composición que proporcione al menos 0,1, 0,5, 1, 1,5, 2, 2,5 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 10 o más unidades de actividad/mg de molécula. En algunos aspectos, el número total de unidades de actividad de la molécula administrada a un sujeto es de al menos 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 110.000, 120.000, 130.000, 140.000, 150.000, 160.000, 170.000, 180.000, 190.000, 200.000, 210.000, 220.000, 230.000 o 250.000, o más unidades. En algunos aspectos, el número total de unidades de actividad de la molécula administrada a un sujeto es, como máximo, de 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 110.000, 120.000, 130.000, 140.000, 150.000, 160.000, 170.000, 180.000, 190.000, 200.000, 210.000, 220.000, 230.000 o 250.000, o más unidades.

En algunos aspectos, se administran al menos aproximadamente 10.000 unidades de actividad a un sujeto, normalizadas por 50 kg de peso corporal. En algunos aspectos, se administran al menos aproximadamente 10.000, 15.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 110.000, 120.000, 130.000, 140.000, 150.000, 160.000, 170.000, 180.000, 190.000, 200.000, 210.000, 220.000, 230.000 o 250.000 unidades o más de actividad de la molécula al sujeto, normalizadas por 50 kg de peso corporal. En algunos aspectos, una dosis terapéuticamente eficaz comprende al menos 5×10^5, 1×10^6, 2×10^6, 3×10^6, 4, 10^6, 5×10^6, 6×10^6, 7×10^6, 8×10^6, 9×10^6, 1×10^7, $1{,}1 \times 10^7$, $1{,}2 \times 10^7$, $1{,}5 \times 10^7$, $1{,}6 \times 10^7$, $1{,}7 \times 10^7$, $1{,}8 \times 10^7$, $1{,}9 \times 10^7$, 2×10^7, $2{,}1 \times 10^7$, o 3×10^7 o más unidades de actividad de la molécula. En algunos aspectos, una dosis terapéuticamente eficaz comprende, como máximo, 5×10^5, 1×10^6, 2×10^6, 3×10^6, 4, 10^6, 5×10^6, 6×10^6, 7×10^6, 8×10^6, 9×10^6, 1×10^7, $1{,}1 \times 10^7$, $1{,}2 \times 10^7$, $1{,}5 \times 10^7$, $1{,}6 \times 10^7$, $1{,}7 \times 10^7$, $1{,}8 \times 10^7$, $1{,}9 \times 10^7$, 2×10^7, $2{,}1 \times 10^7$, o 3×10^7 o más unidades de actividad de la molécula.

En algunos aspectos, una dosis terapéuticamente eficaz es de al menos aproximadamente 10.000, 15.000, 20.000, 22.000, 24.000, 25.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 125.000, 150.000, 200.000 o 500.000 unidades/kg de peso corporal. En algunos aspectos, una dosis terapéuticamente eficaz es como máximo de aproximadamente 10.000, 15.000, 20.000, 22.000, 24.000, 25.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 125.000, 150.000, 200.000 o 500.000 unidades/kg de peso corporal.

En algunos aspectos, la actividad de la molécula administrada a un sujeto es de al menos 10.000, 11.000, 12.000, 13.000, 14,000, 20.000, 21.000, 22.000, 23.000, 24.000, 25.000, 26.000, 27.000, 28.000, 30.000, 32.000, 34.000, 35.000, 36.000, 37.000, 40.000, 45.000 o 50.000, o más U/mg de molécula. En algunos aspectos, la actividad de la molécula administrada a un sujeto es como máximo de 10.000, 11.000, 12.000, 13.000, 14,000, 20.000, 21.000, 22.000, 23.000, 24.000, 25.000, 26.000, 27.000, 28.000, 30.000, 32.000, 34.000, 35.000, 36.000, 37.000, 40.000, 45.000 o 50.000, o más U/mg de molécula.

Mediciones farmacocinéticas y farmacodinámicas

Los datos farmacocinéticos y farmacodinámicos se pueden obtener mediante diversas técnicas experimentales. Los componentes apropiados del perfil farmacocinético y farmacodinámico que describen una determinada composición pueden variar debido a variaciones en el metabolismo del fármaco en los sujetos humanos. Los perfiles farmacocinéticos y farmacodinámicos pueden basarse en la determinación de los parámetros medios de un grupo de sujetos. El grupo de sujetos incluye cualquier número razonable de sujetos adecuados para determinar una media representativa, por ejemplo, 5 sujetos, 10 sujetos, 15 sujetos, 20 sujetos, 25 sujetos, 30 sujetos, 35 sujetos o más. La media se puede determinar calculando el promedio de las mediciones de todos los sujetos para cada parámetro medido. Se puede modular una dosis para lograr un perfil farmacocinético o farmacodinámico deseado, tal como un perfil sanguíneo deseado o eficaz, como se describe en el presente documento.

Los parámetros farmacocinéticos pueden ser cualquier parámetro adecuado para describir una molécula. Por ejemplo, la $C_{máx}$ puede ser, por ejemplo, de no menos de aproximadamente 25 ng/ml; no menos de aproximadamente 50 ng/ml; no menos de aproximadamente 75 ng/ml; no menos de aproximadamente 100 ng/ml; no menos de aproximadamente 200 ng/ml; no menos de aproximadamente 300 ng/ml; no menos de aproximadamente 400 ng/ml; no menos de aproximadamente 500 ng/ml; no menos de aproximadamente 600 ng/ml; no menos de aproximadamente 700 ng/ml; no menos de aproximadamente 800 ng/ml; no menos de aproximadamente 900 ng/ml; no menos de aproximadamente 1000 ng/ml; no menos de aproximadamente 1250 ng/ml; no menos de aproximadamente 1500 ng/ml; no menos de aproximadamente 1750 ng/ml; no menos de aproximadamente 2000 ng/ml; o cualquier otra $C_{máx}$ apropiada para describir un perfil farmacocinético de una molécula descrita en el presente documento.

El $T_{máx}$ de una molécula descrita en el presente documento puede ser, por ejemplo, de no más de aproximadamente 0,5 horas, no más de aproximadamente 1 horas, no más de aproximadamente 1,5 horas, no más de aproximadamente 2 horas, no más de aproximadamente 2,5 horas, no más de aproximadamente 3 horas, no más de aproximadamente 3,5 horas, no más de aproximadamente 4 horas, no más de aproximadamente 4,5 horas, no más de aproximadamente 5 horas, o cualquier otro $T_{máx}$ apropiado para describir un perfil farmacocinético de una molécula descrita en el presente documento.

La $ABC_{(0\text{-}inf)}$ de una molécula descrita en el presente documento puede ser, por ejemplo, de no menos de aproximadamente 50 ng•h/ml, no menos de aproximadamente 100 ng/h/ml, no menos de aproximadamente 150 ng/h/ml, no menos de aproximadamente 200 ng•h/ml, no menos de aproximadamente 250 ng/h/ml, no menos de aproximadamente 300 ng/h/ml, no menos de aproximadamente 350 ng/h/ml, no menos de aproximadamente 400 ng/h/ml, no menos de aproximadamente 450 ng/h/ml, no menos de aproximadamente 500 ng/h/ml, no menos de aproximadamente 600 ng/h/ml, no menos de aproximadamente 700 ng/h/ml, no menos de aproximadamente 800 ng/h/ml, no menos de aproximadamente 900 ng/h/ml, no menos de aproximadamente 1000 ng•h/ml, no menos de aproximadamente 1250 ng/h/ml, no menos de aproximadamente 1500 ng/h/ml, no menos de aproximadamente 1750 ng/h/ml, no menos de aproximadamente 2000 ng/h/ml, no menos de aproximadamente 2500 ng/h/ml, no menos de aproximadamente 3000 ng/h/ml, no menos de aproximadamente 3500 ng/h/ml, no menos de aproximadamente 4000 ng/h/ml, no menos de aproximadamente 5000 ng/h/ml, no menos de aproximadamente 6000 ng/h/ml, no menos de aproximadamente 7000 ng/h/ml, no menos de aproximadamente 8000 ng/h/ml, no menos de aproximadamente 9000 ng/h/ml, no menos de aproximadamente 10.000 ng/h/ml, o cualquier otro valor de $ABC_{(0\text{-}inf)}$ apropiado para describir un perfil farmacocinético de una molécula descrita en el presente documento.

La concentración en plasma de una molécula descrita en el presente documento aproximadamente una hora después de la administración puede ser, por ejemplo, de no menos de aproximadamente 25 ng/ml, no menos de aproximadamente 50 ng/ml, no menos de aproximadamente 75 ng/ml, no menos de aproximadamente 100 ng/ml, no menos de aproximadamente 150 ng/ml, no menos de aproximadamente 200 ng/ml, no menos de aproximadamente 300 ng/ml, no menos de aproximadamente 400 ng/ml, no menos de aproximadamente 500 ng/ml, no menos de aproximadamente 600 ng/ml, no menos de aproximadamente 700 ng/ml, no menos de aproximadamente 800 ng/ml, no menos de aproximadamente 900 ng/ml, no menos de aproximadamente 1000 ng/ml, no menos de aproximadamente 1200 ng/ml, o cualquier otra concentración en plasma de una molécula descrita en el presente documento.

Los parámetros farmacodinámicos pueden ser cualquier parámetro adecuado para describir las composiciones farmacéuticas de la descripción. Por ejemplo, el perfil farmacodinámico puede presentar disminuciones en los factores asociados con la inflamación después de, por ejemplo, aproximadamente 2 horas, aproximadamente 4 horas, aproximadamente 8 horas, aproximadamente 12 horas o aproximadamente 24 horas.

Sales farmacéuticamente aceptables

La divulgación proporciona el uso de sales farmacéuticamente aceptables de cualquier molécula descrita en el presente documento. Las sales farmacéuticamente aceptables pueden incluir, por ejemplo, sales de adición de ácido y sales de adición de base. El ácido que se añade al compuesto para formar una sal de adición de ácido puede ser un ácido orgánico o un ácido inorgánico. Una base que se añade al compuesto para formar una sal de adición de base puede ser una base orgánica o una base inorgánica. En algunos aspectos, una sal farmacéuticamente aceptable es una sal metálica. En algunos aspectos, una sal farmacéuticamente aceptable es una sal de amonio.

Las sales metálicas pueden surgir de la adición de una base inorgánica a un compuesto de la invención. La base inorgánica consiste en un catión metálico emparejado con un contraión básico, tal como, por ejemplo, hidróxido, carbonato, bicarbonato o fosfato. El metal puede ser un metal alcalino, metal alcalinotérreo, metal de transición o metal del grupo principal. En algunos aspectos, el metal es litio, sodio, potasio, cesio, cerio, magnesio, manganeso, hierro, calcio, estroncio, cobalto, titanio, aluminio, cobre, cadmio o cinc.

En algunos aspectos, una sal de metal es una sal de litio, una sal de sodio, una sal de potasio, una sal de cesio, una sal de cerio, una sal de magnesio, una sal de manganeso, una sal de hierro, una sal de calcio, una sal de estroncio, una sal de cobalto, una sal de titanio, una sal de aluminio, una sal de cobre, una sal de cadmio o una sal de cinc, o cualquier combinación de las mismas.

Las sales de amonio pueden surgir de la adición de amoniaco o una amina orgánica a un compuesto de la invención. En algunos aspectos, la amina orgánica es trietilamina, diisopropilamina, etanolamina, dietanolamina, trietanolamina, morfolina, *N*-metilmorfolina, piperidina, *N*-metilpiperidina, *N*-etilpiperidina, dibencilamina, piperazina, piridina, pirrazol, pipirrazol, imidazol, pirazina o pipirazina, o cualquier combinación de las mismas.

En algunos aspectos, una sal de amonio es una sal de trietilamina, una sal de diisopropilamina, una sal de etanolamina, una sal de dietanolamina, una sal de trietanolamina, una sal de morfolina, una sal de *N*-metilmorfolina, una sal de piperidina, una sal de *N*-metilpiperidina, una sal de *N*-etilpiperidina, una sal de dibencilamina, una sal de piperazina, una sal de piridina, una sal de pirrazol, una sal de pipirrazol, una sal de imidazol, una sal de pirazina o una sal de pipirazina, o cualquier combinación de las mismas.

Las sales de adición de ácido pueden surgir de la adición de un ácido a una molécula de la divulgación. En algunos aspectos, el ácido es orgánico. En algunos aspectos, el ácido es inorgánico. En algunos aspectos, el ácido es ácido clorhídrico, ácido bromhídrico, ácido yodhídrico, ácido nítrico, ácido nitroso, ácido sulfúrico, ácido sulfuroso, un ácido fosfórico, ácido isonicotínico, ácido láctico, ácido salicílico, ácido tartárico, ácido ascórbico, ácido gentisínico, ácido glucónico, ácido glucarónico, ácido sacarínico, ácido fórmico, ácido benzoico, ácido glutámico, ácido pantoténico,

ácido acético, ácido propiónico, ácido butírico, ácido fumárico, ácido succínico, ácido metanosulfónico, ácido etanosulfónico, ácido bencenosulfónico, ácido *p*-toluenosulfónico, ácido cítrico, ácido oxálico o ácido maleico, o cualquier combinación de los mismos.

En algunos aspectos, la sal es una sal de clorhidrato, una sal de bromhidrato, una sal de yodhidrato, una sal de nitrato, una sal de nitrito, una sal de sulfato, una sal de sulfito, una sal de fosfato, sal de isonicotinato, una sal de lactato, una sal de salicilato, una sal de tartrato, una sal de ascorbato, una sal de gentisinato, una sal de gluconato, una sal de glucaronato, una sal de sacarato, una sal de formiato, una sal de benzoato, una sal de glutamato, una sal de pantotenato, una sal de acetato, una sal de propionato, una sal de butirato, una sal de fumarato, una sal succinato, una sal de metanosulfonato, una sal de etanosulfonato, una sal de bencenosulfonato, una sal de *p*-toluenosulfonato, una sal de citrato, una sal de oxalato o una sal de maleato, o cualquier combinación de las mismas.

Polipéptidos dirigidos a un sitio diseñados

Visión general

La divulgación describe métodos, composiciones, sistemas y/o kits para modificar polipéptidos dirigidos a un sitio (por ejemplo, Cas9, Csy4, Cas5, Cas6, Argonauta, etc.) y/o enzimas relacionadas. Las modificaciones pueden incluir cualquier modificación covalente o no covalente en los polipéptidos dirigidos a un sitio. En algunos casos, esto puede incluir modificaciones químicas en una o más regiones del polipéptido dirigido a un sitio. En algunos casos, las modificaciones pueden incluir sustituciones de aminoácidos conservativas o no conservativas del polipéptido dirigido a un sitio. En algunos casos, las modificaciones pueden incluir la adición, eliminación o sustitución de cualquier parte del polipéptido dirigido a un sitio con aminoácidos, péptidos o dominios que no se encuentran en el polipéptido nativo dirigido a un sitio. En algunos casos, uno o más dominios no nativos se pueden añadir, eliminar o sustituir en el polipéptido dirigido a un sitio. En algunos casos, el polipéptido dirigido a un sitio puede existir como una proteína de fusión.

En algunos casos, la presente divulgación proporciona el diseño a de polipéptidos dirigidos a un sitio para reconocer una secuencia de ácido nucleico diana deseada con especificidad y/o actividad enzimática deseadas. Las modificaciones en un polipéptido dirigido a un sitio se pueden realizar mediante el diseño de proteínas. El diseño de proteínas puede incluir la fusión de dominios funcionales a dicho polipéptido dirigido a un sitio diseñado que puede usarse para modificar el estado funcional del polipéptido dirigido a un sitio global o la secuencia de ácido nucleico diana real de un locus celular endógeno. El polipéptido dirigido a un sitio de la divulgación puede usarse para regular la expresión de genes endógenos, tanto a través de la activación como de la represión de la transcripción de genes endógenos.

Las fusiones de polipéptidos dirigidos al sitio también se pueden unir a otros dominios reguladores o funcionales, por ejemplo, nucleasas, transposasas o metilasas, para modificar secuencias cromosómicas endógenas. En algunos casos, el polipéptido dirigido a un sitio puede estar unido a al menos uno o más dominios reguladores, descritos en el presente documento. Los ejemplos no limitantes de dominios reguladores o funcionales incluyen dominios represores o activadores de factores de transcripción tales como KRAB y VP 16, dominios co-represores y co-activadores, metil transferasas de ADN, histona acetiltransferasas, histona desacetolasas y dominios de escisión de ADN tal como el dominio de escisión de la endonucleasa FokI.

En algunos casos, uno o más dominios específicos, regiones o elementos estructurales del polipéptido dirigido a un sitio pueden modificarse juntos. Pueden producirse modificaciones en el polipéptido dirigido a un sitio, pero no se limitan a los elementos del polipéptido dirigido a un sitio, tales como regiones que reconocen o se unen al motivo adyacente al espaciador (PAM), y/o regiones que se unen a o reconocen el ácido nucleico dirigido a ácido nucleico. Dichos elementos de unión o reconocimiento pueden incluir una hélice de puente conservada, región altamente básica, región de N-terminal, región C-terminal, motivos RuvC (por ejemplo, dominios nucleasa de tipo RuvC y/o RuvC) y uno o más dominios nucleasa, tales como HNH y/o dominios de tipo HNH. Se pueden realizar modificaciones en dominios adicionales, elementos estructurales, secuencia o aminoácidos dentro del polipéptido dirigido a un sitio.

Se pueden realizar modificaciones en una o más regiones del polipéptido dirigido a un sitio para alterar diversas propiedades del polipéptido dirigido a un sitio. En algunos casos, las modificaciones pueden alterar el reconocimiento de unión para ciertas secuencias diana de ácido nucleico. Esto puede incluir, pero sin limitación, aumentar la afinidad y/o especificidad de unión a ciertas secuencias o preferentemente la dirección de ciertas secuencias de ácido nucleico/elementos de reconocimiento. En algunos casos, se pueden usar modificaciones para alterar la función de la nucleasa nativa. En algunos casos, las modificaciones realizadas en el polipéptido dirigido a un sitio pueden alterar la especificidad hacia PAM, la especificidad de ARNcrtra, la especificidad de ARNcr o la especificidad para elementos de ácido nucleico adicionales, tales como un ácido nucleico dirigido a ácido nucleico.

En el presente documento, también se describen composiciones y métodos que incluyen proteínas de fusión que comprenden un polipéptido dirigido a un sitio (por ejemplo, Cas9) y uno o más dominios o regiones diseñados para la edición genómica (por ejemplo, escisión de genes; alteración de genes, por ejemplo, mediante escisión seguida de

inserción (inserción física o inserción mediante reparación dirigida por homología) de una secuencia exógena y/o escisión seguida de NHEJ; inactivación parcial o total de uno o más genes; generación de alelos con estados funcionales alterados de genes endógenos, inserción de elementos reguladores; etc.) y alteraciones del genoma que se portan a la línea germinal. También se desvelan métodos de fabricación y uso de estas composiciones (es decir, reactivos), por ejemplo, para editar (es decir, alterar) uno o más genes de una célula diana. Así pues, los métodos y las composiciones descritos en el presente documento proporcionan métodos altamente eficaces para la alteración genética dirigida (por ejemplo, inserción) y/o desactivación (parcial o total) de uno o más genes y/o para la mutación aleatoria de la secuencia de cualquier alelo diana y, por tanto, permiten la generación de modelos animales de enfermedades humanas. Un experto en la materia reconocerá que, aunque la expresión "diseño del genoma" o "edición genómica" se suele usar para describir los métodos del presente documento, los métodos y las composiciones descritos en el presente documento también pueden usarse para alterar cualquier ácido nucleico diana que puede no estar, estrictamente hablando, en el genoma de una célula (por ejemplo, se puede usar en un ácido nucleico sintético, un plásmido, un vector, un ácido nucleico vírico, un ácido nucleico recombinante, etc.).

Los métodos y las composiciones descritos en el presente documento permiten nuevas aplicaciones terapéuticas, (por ejemplo, prevención y/o tratamiento de: enfermedades genéticas, cáncer, Infecciones fúngicas, protozoarias, bacterianas y víricas, isquemia, enfermedad vascular, artritis, trastornos inmunológicos, etc.), diagnósticos novedosos (por ejemplo, predicción y/o diagnóstico de una afección), así como proporcionar herramientas de investigación (por ejemplo kits, ensayos de genómica funcional, y generación de estirpes celulares diseñadas y modelos animales diseñados para investigación y detección de fármacos), y medios para desarrollar plantas con fenotipos alterados, incluyendo, pero sin limitación, aumento de la resistencia a enfermedades y alteración de las características de maduración de la fruta, composición de azúcar y aceite, rendimiento y color. Los métodos y las composiciones descritos en el presente documento permiten nuevos estudios epigenéticos.

Modificaciones y diseño de proteínas

Alteraciones de aminoácidos

Los polipéptidos dirigidos a un sitio, como se desvela en el presente documento, pueden modificarse. La modificación puede comprender modificaciones en un aminoácido del polipéptido dirigido a un sitio. Las modificaciones pueden alterar la secuencia de aminoácidos primaria y/o la estructura de aminoácidos secundaria, terciaria y cuaternaria. En algunos casos, algunas secuencias de aminoácidos de un polipéptido dirigido a un sitio pueden variarse sin un efecto significativo sobre la estructura ni la función de la proteína. El tipo de mutación puede ser completamente irrelevante si la alteración ocurre en algunas regiones (por ejemplo, una región no fundamental) de la proteína. En algunos casos, dependiendo de la ubicación del reemplazo, la mutación puede no tener un efecto importante en las propiedades biológicas de la variante resultante. Por ejemplo, las propiedades y funciones de las variantes de Cas9 pueden ser del mismo tipo que Cas9 de tipo silvestre. En algunos casos, la mutación puede tener un impacto crítico en la estructura y/o función del polipéptido dirigido a un sitio.

La ubicación de dónde modificar un polipéptido dirigido a un sitio (por ejemplo, una variante de Cas9) puede determinarse usando la alineación de secuencias y/o estructuras. La alineación de secuencias puede identificar regiones de un polipéptido que sean similares y/o no similares (por ejemplo, conservadas, no conservadas, hidrófobas, hidrófilas, etc.). En algunos casos, una región de la secuencia de interés que es similar a otras secuencias es adecuada para la modificación. En algunos casos, una región de la secuencia de interés que es diferente de otras secuencias es adecuada para la modificación. Por ejemplo, la alineación de secuencias se puede realizar mediante búsqueda en bases de datos, alineación por pares, alineación de múltiples secuencias, análisis genómico, búsqueda de motivos, evaluación comparativa y/o programas tales como BLAST, CS-BLAST, HHPRED, psi-BLAST, LALIGN, PyMOL y SEQALN. La alineación estructural puede realizarse mediante programas tales como Dali, PHYRE, Chimera, COOT, O y PyMOL. La alineación se puede realizar mediante búsqueda en bases de datos, alineación por pares, alineación de múltiples secuencias, análisis genómico, búsqueda de motivos o evaluación comparativa, o cualquier combinación de los mismos.

Un polipéptido dirigido a un sitio puede modificarse para aumentar la especificidad de unión a un ácido nucleico dirigido a ácido nucleico y/o un ácido nucleico diana. Un polipéptido dirigido a un sitio puede modificarse para aumentar la unión a regiones específicas de un ácido nucleico dirigido a ácido nucleico (por ejemplo, la extensión del espaciador, el espaciador, la repetición CRISPR mínima, la secuencia mínima de ARNcrtra, la secuencia de ARNcrtra de 3', la extensión de ARNcrtra) y/o un ácido nucleico diana.

En algunos casos, la modificación puede comprender una modificación conservativa. Un cambio conservativo de aminoácidos puede implicar la sustitución de uno de una familia de aminoácidos que están relacionados en sus cadenas laterales (por ejemplo, cisteína/serina).

En algunos casos, los cambios de aminoácidos en la proteína Cas9 descritos en el presente documento son cambios de aminoácidos no conservativos (es decir, sustituciones de aminoácidos con carga o sin carga similares). Un cambio de aminoácido no conservativo puede implicar la sustitución de uno de una familia de aminoácidos que pueden no estar relacionados en sus cadenas laterales o una sustitución que altere la actividad biológica del polipéptido dirigido

a un sitio.

La mutación de los aminoácidos también puede cambiar la selectividad de la unión a un ácido nucleico diana. La mutación puede dar lugar a un cambio que puede comprender un cambio en la constante de disociación (K_d) de la unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico diana. El cambio en la K_d de unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico diana puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a la K_d de la unión entre un polipéptido no mutado dirigido a un sitio y una ácido nucleico diana. El cambio en la K_d de la unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico diana puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a la K_d de la unión entre un polipéptido no mutado dirigido a un sitio y un ácido nucleico diana.

La mutación puede dar lugar a un cambio que puede comprender un cambio en la K_d de unión entre un polipéptido mutado dirigido a un sitio y un motivo PAM. El cambio en la K_d de unión entre un polipéptido mutado dirigido a un sitio y un motivo PAM puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a la K_d de unión entre un polipéptido no mutado dirigido a un sitio y un motivo PAM. El cambio en la K_d de la unión entre un polipéptido mutado dirigido a un sitio y un motivo PAM puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a la K_d de la unión entre un polipéptido no mutado dirigido a un sitio y un motivo PAM.

La mutación puede dar lugar a un cambio que puede comprender un cambio en la K_d de unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico. El cambio en la K_d de unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a la K_d de unión entre un polipéptido no mutado dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico. El cambio en la K_d de la unión entre un polipéptido mutado dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a la K_d de la unión entre un polipéptido no mutado dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico.

La mutación de un polipéptido dirigido a un sitio también puede cambiar la cinética de la acción enzimática del polipéptido dirigido a un sitio. La mutación puede dar lugar a un cambio que puede comprender un cambio en la K_m del polipéptido mutado dirigido a un sitio. El cambio en la K_m del polipéptido mutado dirigido a un sitio puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a la K_m de un polipéptido no mutado dirigido a un sitio. El cambio en la K_m de un polipéptido mutado dirigido a un sitio puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a la K_m de un polipéptido no mutado dirigido a un sitio.

La mutación de un polipéptido dirigido a un sitio puede producir un cambio que puede comprender un cambio en el recambio del polipéptido dirigido a un sitio. El cambio en el recambio del polipéptido mutado dirigido a un sitio puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior al recambio de un polipéptido no mutado dirigido a un sitio. El cambio en el recambio de un polipéptido mutado dirigido a un sitio puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior al recambio de un polipéptido no mutado dirigido a un sitio.

La mutación puede producir un cambio que puede comprender un cambio en ΔG de la acción enzimática del polipéptido dirigido a un sitio. El cambio en ΔG del polipéptido mutado dirigido a un sitio puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a ΔG de un polipéptido no mutado dirigido a un sitio. El cambio en el recambio de un polipéptido mutado dirigido a un sitio puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a ΔG de un polipéptido no mutado dirigido a un sitio.

La mutación puede producir un cambio que puede comprender un cambio en $V_{máx}$ de la acción enzimática del polipéptido dirigido a un sitio. El cambio en $V_{máx}$ del polipéptido mutado dirigido a un sitio puede ser más de 1000 veces, más de 500 veces, más de 100 veces, más de 50 veces, más de 25 veces, más de 10 veces, más de 5 veces, más de 4 veces, más de 3 veces, más de 2 veces superior o inferior a $V_{máx}$ de un polipéptido no mutado dirigido a un sitio.

El cambio en el recambio de un polipéptido mutado dirigido a un sitio puede ser menos de 1000 veces, menos de 500 veces, menos de 100 veces, menos de 50 veces, menos de 25 veces, menos de 10 veces, menos de 5 veces, menos de 4 veces, menos de 3 veces, menos de 2 veces superior o inferior a $V_{máx}$ de un polipéptido no mutado dirigido a un sitio.

La mutación puede producir un cambio que puede comprender un cambio en cualquier parámetro cinético del polipéptido dirigido a un sitio. La mutación puede producir un cambio que puede comprender un cambio en cualquier parámetro termodinámico del polipéptido dirigido a un sitio. La mutación puede producir un cambio que puede comprender un cambio en la carga superficial, el área superficial enterrada y/o la cinética de plegamiento del polipéptido dirigido a un sitio y/o la acción enzimática del polipéptido dirigido a un sitio.

Los aminoácidos del polipéptido dirigido a un sitio de la presente divulgación que son esenciales para la función pueden identificarse mediante métodos tales como mutagénesis dirigida, mutagénesis de barrido de alanina, análisis de estructura de proteínas, resonancia magnética nuclear, marcaje de fotoafinidad y tomografía de electrones, cribado de alto rendimiento, ELISA, ensayos bioquímicos, ensayos de unión, ensayos de escisión (por ejemplo, ensayo Surveyor), ensayos de indicadores y similares.

Otras alteraciones de aminoácidos también pueden incluir aminoácidos con formas glicosiladas, conjugados agregativos con otras moléculas y conjugados covalentes con fracciones químicas no relacionadas (por ejemplo, moléculas pegiladas). Las variantes covalentes pueden prepararse uniendo funcionalidades a grupos que se encuentran en la cadena de aminoácidos o en el resto N o C-terminal. En algunos casos, los polipéptidos mutados dirigidos a un sitio también pueden incluir variantes alélicas y variantes de especies.

Se pueden diseñar truncamientos de regiones que no afecten a la actividad funcional de las proteínas Cas9. Se pueden diseñar truncamientos de regiones que afecten a la actividad funcional de la proteína Cas9. Un truncamiento puede comprender un truncamiento de menos de 5, menos de 10, menos de 15, menos de 20, menos de 25, menos de 30, menos de 35, menos de 40, menos de 45, menos de 50, menos de 60, menos de 70, menos de 80, menos de 90, menos de 100 o más aminoácidos. Un truncamiento puede comprender un truncamiento de más de 5, más de 10, más de 15, más de 20, más de 25, más de 30, más de 35, más de 40, más de 45, más de 50, más de 60, más de 70, más de 80, más de 90, más de 100 o más aminoácidos. Un truncamiento puede comprender un truncamiento del aproximadamente 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % del polipéptido dirigido a un sitio.

Se pueden diseñar eliminaciones de regiones que no afecten a la actividad funcional de las proteínas Cas9. Se pueden diseñar eliminaciones de regiones que afecten a la actividad funcional de la proteína Cas9. Una eliminación puede comprender una eliminación de menos de 5, menos de 10, menos de 15, menos de 20, menos de 25, menos de 30, menos de 35, menos de 40, menos de 45, menos de 50, menos de 60, menos de 70, menos de 80, menos de 90, menos de 100 o más aminoácidos. Una eliminación puede comprender un eliminación de más de 5, más de 10, más de 15, más de 20, más de 25, más de 30, más de 35, más de 40, más de 45, más de 50, más de 60, más de 70, más de 80, más de 90, más de 100 o más aminoácidos. Una eliminación puede comprender una eliminación del aproximadamente 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % del polipéptido dirigido a un sitio. Una eliminación puede ocurrir en el extremo N, el extremo C o en cualquier región de la cadena polipeptídica.

Cribados

La divulgación proporciona métodos de diseño de un polipéptido dirigido a un sitio. Se pueden usar cribados para diseñar un polipéptido dirigido a un sitio. Por ejemplo, se puede configurar un cribado para detectar el efecto de las mutaciones en una región del polipéptido dirigido a un sitio. Por ejemplo, se puede configurar un cribado para probar modificaciones del parche altamente básico en la afinidad por la estructura de ARN (por ejemplo, la estructura del ácido nucleico dirigida a ácido nucleico) o la capacidad de procesamiento (por ejemplo, la escisión de ácido nucleico diana). Los métodos de cribado ilustrativos pueden incluir, pero sin limitación, métodos de clasificación celular, presentación en ARNm, presentación en fagos y evolución dirigida.

Fusiones

En algunos casos, el polipéptido dirigido a un sitio se modifica de modo que comprenda una secuencia no nativa (es decir, el polipéptido tenga una modificación que lo altera con respecto al alelo o de la secuencia de la que se deriva) (por ejemplo, el polipéptido puede denominarse fusión). La secuencia no nativa también puede incluir una o más proteínas adicionales, dominios de proteínas, subdominios o polipéptidos. Por ejemplo, Cas9 puede fusionarse con cualquier proteína y/o dominio de unión a ácido nucleico normativo adicional adecuado, incluyendo, pero sin limitación, dominios de factor de transcripción, dominios nucleasa, dominios de polimerización de ácido nucleico. La secuencia no nativa puede comprender una secuencia de Cas9 y/o un homólogo de Cas9.

La secuencia no nativa puede conferir nuevas funciones a la proteína de fusión. Estas funciones pueden incluir, por ejemplo, escisión del ADN, metilación del ADN, daño del ADN, reparación del ADN, modificación de un polipéptido

diana asociado con el ADN diana (por ejemplo, una histona, una proteína de unión al ADN, etc.), lo que lleva a, por ejemplo, la metilación de histonas, acetilación de histonas, ubiquitinación de histonas y similares. Otras funciones conferidas por una proteína de fusión pueden incluir actividad de metiltransferasa, actividad de desmetilasa, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de despurinación, actividad de oxidación, actividad de formación de dímeros de pirimidina, actividad integrasa, actividad de transposasa, actividad de recombinasa, actividad polimerasa, actividad de ligasa, actividad de helicasa, actividad de fotoliasa o actividad de glucosilasa, actividad de acetiltransferasa, actividad de desacetilasa, actividad de quinasa, actividad fosfatasa, actividad de ubiquitina ligasa, actividad desubiquitinante, actividad de adenilación, actividad de desadenilación, actividad de SUMOilación, actividad deSUMOilación, actividad de ribosilación, actividad de desribosilación, actividad de miristoilación, actividad de remodelación, actividad de proteasa, actividad de oxidorreductasa, actividad de transferasa, actividad de hidrolasa, actividad de liasa, actividad de isomerasa, actividad de sintasa, actividad de sintetasa y actividad de desmiristoilación, o cualquier combinación de las mismas.

Modificaciones en la hélice puente

En algunos casos, se puede modificar la región de la hélice puente de Cas9 (por ejemplo, para alterar la especificidad hacia PAM). En algunos casos, la hélice puente puede compartir homología con la hélice puente identificada en la proteína Cas9 de *S.pyogenes* (restos 551-566). En algunos casos, la hélice puente puede compartir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 551-556 de la hélice puente de Cas9 de *S. pyogenes.* En algunos casos, la hélice puente puede compartir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 551-556 de la hélice puente de Cas9 de *S. pyogenes.*

En algunos casos, las modificaciones en la hélice puente pueden incluir, pero sin limitación, modificaciones de aminoácidos individuales, como se describe en el presente documento. En algunos casos, la modificación de la hélice puente puede incluir, pero sin limitación, inserciones, eliminaciones o sustituciones de aminoácidos o polipéptidos individuales, tales como otros elementos proteicos (por ejemplo, dominios, motivos estructurales, proteínas).

Las modificaciones pueden incluir modificaciones de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos de la hélice puente. Las modificaciones pueden incluir modificaciones, como máximo, en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos de la hélice puente. Las modificaciones también pueden incluir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la hélice puente. Las modificaciones también pueden incluir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la hélice puente.

En algunos casos, las modificaciones en las secuencias de la hélice puente del polipéptido dirigido a un sitio pueden incluir determinados motivos estructurales del polipéptido, incluyendo, aunque sin limitación, la hélice alfa, cadena beta, lámina beta, hélice 310, hélice pi, motivo de polipropolina I, motivo de polipropolina II, motivo de poliprolina III, vuelta beta, alfa-vuelta-alfa, o acodamientos o bisagras de hélice. Por ejemplo, las sustituciones en la hélice puente del polipéptido dirigido a un sitio pueden incluir la sustitución o adición con uno o más restos de aminoácidos de prolina. La inserción de restos de prolina puede introducir acodamientos en la hélice puente que pueden alterar la especificidad de unión de la hélice puente por el PAM. En otro ejemplo, la sustitución o adición puede incluir uno o más restos de aminoácidos de glicina. La inserción o sustitución de restos de glicina puede introducir una mayor flexibilidad en la hélice puente o "bisagras" que también pueden alterar la especificidad de unión de la hélice puente por el PAM. La alteración de la especificidad de unión puede o no afectar a la actividad enzimática de la proteína Cas9.

En algunos casos, las modificaciones en las secuencias de la hélice puente del polipéptido dirigido a un sitio pueden incluir la eliminación de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la hélice puente. En algunos casos, las modificaciones en las secuencias de la hélice puente del polipéptido dirigido a un sitio pueden incluir la eliminación de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la hélice puente.

En algunos casos, las modificaciones en las secuencias de la hélice puente del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de una hélice puente del polipéptido dirigido a un sitio homóloga. En algunos casos, las modificaciones en las secuencias de la hélice puente del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de una hélice puente del polipéptido dirigido a un sitio homóloga.

Por ejemplo, se pueden derivar hélices puente de Cas9 no nativas de cualquier organismo adecuado. En algunos casos, la proteína Cas9 y la hélice puente pueden derivarse de organismos procariotas, incluyendo, pero sin limitación, arqueas, bacterias, protistas (por ejemplo,, *E. coli, S. pyogenes,* S. *thermophilus, P.furiosus*, etc.).

Por ejemplo, la hélice puente de la enzima Cas9 de *S. pyogenes* puede sustituirse o insertarse con la hélice puente, o fragmento de la misma, derivada de otra enzima Cas9 de una especie diferente.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y una hélice puente modificada.

Modificaciones del parche altamente básico

La unión y especificidad hacia PAM también pueden verse afectadas por regiones adicionales dentro de la proteína Cas9. En algunos casos, un parche altamente básico o una región altamente básica, que comprende restos de aminoácidos básicos adyacentes a un sitio de unión a PAM, también puede modificarse para alterar la especificidad hacia PAM. En algunos casos, el parche o la región altamente básico puede compartir homología con el parche altamente básico identificado en la Cas9 de *S. pyogenes* contenida dentro de la región N-terminal, o los restos de aminoácidos 1-270 del Cas9 de *S. pyogenes*. En algunos casos, el parche altamente básico puede compartir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con el parche altamente básico de Cas9 de *S. pyogenes*. En algunos casos, el parche altamente básico puede compartir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con el parche altamente básico de Cas9 de *S. pyogenes*.

En algunos casos, las modificaciones en el parche altamente básico pueden incluir, pero sin limitación, modificaciones de aminoácidos individuales, como se describe en el presente documento. En algunos casos, la modificación en el parche altamente básico puede incluir, pero sin limitación, inserciones, eliminaciones o sustituciones de aminoácidos o polipéptidos individuales, tales como otros elementos proteicos (por ejemplo, dominios, motivos estructurales, proteínas).

Las modificaciones pueden incluir modificaciones de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37,38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 o más aminoácidos del parche altamente básico. Las modificaciones pueden incluir modificaciones de, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 o más aminoácidos del parche altamente básico. Las modificaciones también pueden incluir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del parche altamente básico. Las modificaciones también pueden incluir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del parche altamente básico.

En algunos casos, las modificaciones de la secuencia del parche altamente básico del polipéptido dirigido a un sitio pueden incluir determinados motivos estructurales del polipéptido, incluyendo, aunque sin limitación, la hélice alfa, cadena beta, lámina beta, hélice 310, hélice pi, motivo de polipropolina I, motivo de polipropolina II, motivo de poliprolina III, vuelta beta, alfa-vuelta-alfa, o acodamientos o bisagras de hélice.

Las sustituciones del parche altamente básico del polipéptido dirigido a un sitio pueden incluir la sustitución o adición con uno o más restos de aminoácidos ácidos. La inserción de restos ácidos puede disminuir la carga básica general de esta área del polipéptido dirigido a un sitio y puede alterar la especificidad de unión del parche altamente básico hacia el PAM. En otro ejemplo, la sustitución o adición puede incluir uno o más restos de aminoácidos básicos. La inserción o sustitución de restos básicos puede aumentar el área de carga o la fuerza iónica de interacción entre el polipéptido y el ácido nucleico, y también puede alterar la especificidad de unión del parche altamente básico hacia el PAM. La alteración de la especificidad de unión puede o no afectar a la actividad enzimática del polipéptido dirigido a un sitio.

En algunos casos, las modificaciones de las secuencias de parche altamente básico del polipéptido dirigido a un sitio pueden incluir la eliminación de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del parche altamente básico. En algunos casos, las modificaciones de las secuencias de parches altamente básicos del polipéptido dirigido a un sitio pueden incluir la eliminación de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del parche altamente básico.

En algunos casos, las modificaciones de las secuencias de parches altamente básicos del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de un parche altamente básico de Cas9 homólogo. En algunos casos, las modificaciones de las secuencias de parches altamente básicos del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de un parche altamente básico de Cas9 homólogo.

Las secuencias de parches altamente básicos de Cas9 homólogos pueden derivarse de cualquier organismo adecuado. En algunos casos, la proteína Cas9 puede derivarse de organismos procariotas tales como arqueas, bacterias, protistas (por ejemplo,, *E. coli, S. pyogenes, S. thermophilus, P.furiosus,* etc.).Por ejemplo, el parche altamente básico de la enzima Cas9 de *S. pyogenes* puede ser sustituido con o modificado mediante la inserción del parche altamente básico, o fragmento del mismo, derivado de una Cas9 de otra especie.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un parche altamente básico modificado.

Modificaciones del dominio HNH

En algunos casos, el dominio HNH de un polipéptido dirigido a un sitio puede modificarse para alterar la especificidad hacia PAM. En algunos casos, el dominio HNH puede compartir homología con el dominio HNH identificado en el dominio C-terminal de la proteína Cas9 de *S. pyogenes* (restos 860-1100). En algunos casos, el dominio HNH puede compartir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 551-556 del dominio NHN de Cas9 de *S. pyogenes.* En algunos casos, el dominio HNH puede compartir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 860-1100 del dominio HNH de Cas9 de *S. pyogenes*.

En algunos casos, las modificaciones en el dominio HNH pueden incluir, pero sin limitación, modificaciones de aminoácidos individuales, como se describe en el presente documento. En algunos casos, la modificación en el dominio HNH puede incluir, pero sin limitación, inserciones, eliminaciones o sustituciones de aminoácidos o polipéptidos individuales, tales como otros elementos proteicos (por ejemplo, dominios, motivos estructurales, proteínas).

Las modificaciones pueden incluir modificaciones de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos del dominio HNH. Las modificaciones pueden incluir modificaciones, como máximo, en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos del dominio HNH. Las modificaciones también pueden incluir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio HNH. Las modificaciones también pueden incluir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio HNH.

En algunos casos, las modificaciones en las secuencias del dominio HNH del polipéptido dirigido a un sitio pueden incluir determinados motivos estructurales del polipéptido, incluyendo, aunque sin limitación, la hélice alfa, cadena beta, lámina beta, hélice 310, hélice pi, motivo de polipropolina I, motivo de polipropolina II, motivo de poliprolina III, vuelta beta, alfa-vuelta-alfa, o acodamientos o bisagras de hélice.

Las sustituciones del dominio HNH del polipéptido dirigido a un sitio pueden incluir la sustitución o adición con uno o más restos de aminoácidos. En algunos casos, el dominio HNH puede reemplazarse o fusionarse con otros dominios de unión a ácido nucleico adecuados. Un dominio de unión a ácido nucleico puede comprender ARN. Puede haber un solo dominio de unión a ácido nucleico. Los ejemplos de dominios de unión a ácido nucleico pueden incluir, pero sin limitación, un dominio hélice-vuelta-hélice, un dominio de dedo de cinc, un dominio de cremallera de leucina (bZIP), un dominio de hélice alada, un dominio de hélice alada-vuelta-hélice, un dominio hélice-lazo-hélice, un dominio de caja HMG, un dominio Wor3, un dominio de inmunoglobulina, un dominio B3, un dominio TALE, un dominio de dedo de cinc, un dominio de motivo de reconocimiento de ARN, un dominio de motivo de unión a ARN bicatenario, un dominio de unión a ácido nucleico bicatenario, un dominio de unión a ácido nucleico monocatenario, un dominio KH, un dominio PUF, un dominio de caja RGG, un dominio de caja DEAD/DEAH, un dominio PAZ, un dominio Piwi y un dominio de choque frío, un dominio RNAsaH, un dominio HNH, un dominio de tipo RuvC, un dominio RAMP, un dominio Cas5, un dominio Cas6.

En algunos casos, las modificaciones en las secuencias del dominio HNH del polipéptido dirigido a un sitio pueden incluir la eliminación de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio HNH. En algunos casos, las modificaciones en las secuencias del dominio HNH del polipéptido dirigido a un sitio pueden incluir la eliminación de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio HNH.

En algunos casos, las modificaciones en las secuencias del dominio HNH del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de un dominio HNH de Cas9 homólogo. En algunos casos, las modificaciones en las secuencias del dominio HNH del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %,

55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de un dominio HNH de Cas9 homólogo.

Los dominios HNH homólogos de Cas9 pueden derivarse de cualquier organismo adecuado. En algunos casos, la proteína Cas9 puede derivarse de organismos procariotas tales como arqueas, bacterias, protistas (por ejemplo, *E. cob*, *S. pyogenes, S. thermophilus, P.furiosus,* etc.). Por ejemplo, el dominio HNH de la enzima Cas9 de *S. pyogenes* puede ser sustituido con o modificado mediante la inserción del dominio HNH, o fragmento del mismo, derivado de una enzima Cas9 de otra especie. En algunos casos, se puede insertar al menos un dominio HNH homólogo de Cas9 en el dominio HNH. En algunos casos, el al menos un dominio HNH homólogo de Cas9 puede formar una matriz de dominios HNH, que comprende al menos dos dominios HNH. En algunos casos, una matriz de dominio HNH puede comprender al menos un dominio HNH de Cas9 y al menos un segundo dominio HNH.

En algunos casos, la modificación del dominio HNH o de tipo HNH puede comprender la inserción del mismo dominio HNH o de tipo HNH, o de uno similar, en tándem (por ejemplo, adyacente) al dominio HNH de Cas9. El dominio HNH o de tipo HNH puede insertarse en el extremo N y/o extremo C del dominio HNH en Cas9. La inserción de uno o más dominios HNH o de tipo HNH en Cas9 puede ser útil para ampliar la especificidad en un ácido nucleico diana. La inserción de uno o más dominios HNH o de tipo HNH en Cas9 puede ser útil para duplicar la especificidad en un ácido nucleico diana. Por ejemplo, la inserción de uno o más dominios HNH o de tipo HNH puede configurar Cas9 para reconocer un tramo más largo de ácido nucleico diana, reconocer un híbrido de ARN-ADN diferente y/o reconocer un ácido nucleico diana con mayor afinidad de unión.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un dominio HNH modificado.

Modificaciones de los dominios RuvC o de tipo RuvC

En algunos casos, el dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio puede modificarse para alterar la especificidad hacia PAM. En algunos casos, el dominio RuvC o de tipo RuvC puede compartir homología con el dominio RuvC o de tipo RuvC identificado en la proteína Cas9 de *S. pyogenes* (restos 1-270). En algunos casos, el dominio RuvC o de tipo RuvC puede compartir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 551-556 del dominio RuvC o de tipo RuvC de Cas9 de *S. pyogenes.* En algunos casos, el dominio RuvC o de tipo RuvC puede compartir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 1-270 del dominio RuvC o de tipo RuvC de Cas9 de *S. pyogenes*.

En algunos casos, las modificaciones en el dominio RuvC o de tipo RuvC pueden incluir, pero sin limitación, modificaciones de aminoácidos individuales, como se describe en el presente documento. En algunos casos, las modificaciones del dominio RuvC o de tipo RuvC pueden incluir, pero sin limitación, eliminaciones o sustituciones de aminoácidos o polipéptidos individuales, tales como otros elementos proteicos (por ejemplo, dominios, motivos estructurales, proteínas).

Las modificaciones pueden incluir modificaciones de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos del dominio RuvC o de tipo RuvC. Las modificaciones pueden incluir modificaciones, como máximo, en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos del dominio RuvC o de tipo RuvC. Las modificaciones también pueden incluir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC. Las modificaciones también pueden incluir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC.

En algunos casos, las modificaciones de las secuencias de dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir determinados motivos estructurales del polipéptido, incluyendo, aunque sin limitación, la hélice alfa, cadena beta, lámina beta, hélice 310, hélice pi, motivo de polipropolina I, motivo de polipropolina II, motivo de poliprolina III, vuelta beta, alfa-vuelta-alfa, o acodamientos o bisagras de hélice.

Las sustituciones del dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir la sustitución o adición con uno o más restos de aminoácidos. En algunos casos, el dominio RuvC o de tipo RuvC puede reemplazarse o fusionarse con otros dominios de unión a ácido nucleico adecuados. Un dominio de unión a ácido nucleico puede comprender ARN. Puede haber un solo dominio de unión a ácido nucleico. Los ejemplos de dominios de unión a ácido nucleico pueden incluir, pero sin limitación, un dominio hélice-vuelta-hélice, un dominio de dedo de cinc, un dominio de cremallera de leucina (bZIP), un dominio de hélice alada, un dominio de hélice alada-vuelta-hélice, un dominio hélice-lazo-hélice, un dominio de caja HMG, un dominio Wor3, un dominio de inmunoglobulina, un dominio B3, un dominio TALE, un dominio de dedo de cinc, un dominio de motivo de reconocimiento de ARN, un dominio de motivo de unión a ARN bicatenario, un dominio de unión a ácido nucleico bicatenario, un dominio de unión a ácido nucleico monocatenario, un dominio KH, un dominio PUF, un dominio de caja RGG, un dominio de caja DEAD/DEAH, un

dominio PAZ, un dominio Piwi, un dominio de choque frío, un dominio RNAsaH, un dominio HNH, un dominio de tipo RuvC, un dominio RAMP, un dominio Cas5 y un dominio Cas6.

En algunos casos, las modificaciones en las secuencias del dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir la eliminación de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC. En algunos casos, las modificaciones en las secuencias del dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir la eliminación de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC.

En algunos casos, las modificaciones en las secuencias del dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC de Cas9 homólogo. En algunos casos, las modificaciones en las secuencias del dominio RuvC o de tipo RuvC del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del dominio RuvC o de tipo RuvC de Cas9 homólogo.

Los dominios RuvC o de tipo RuvC de Cas9 homólogos se pueden derivar de cualquier organismo adecuado. En algunos casos, la proteína Cas9 puede derivarse de organismos procariotas tales como arqueas, bacterias, protistas (por ejemplo, *E. coli*, *S. pyogenes, S. thermophilus, P.furiosus,* etc.). Por ejemplo, el dominio RuvC o de tipo RuvC de la enzima Cas9 de *S. pyogenes* puede ser sustituido con o modificado mediante la inserción del dominio RuvC o de tipo RuvC, o un fragmento del mismo, derivado de otra enzima Cas9, tal como uno de otra especie.

En algunos casos, la modificación del dominio RuvC o de tipo RuvC puede comprender la inserción del mismo dominio RuvC o de tipo RuvC, o de uno similar, en tándem (por ejemplo, adyacente) al dominio RuvC de Cas9. El dominio RuvC o de tipo RuvC se puede insertar en el extremo N y/o extremo C del dominio RuvC o de tipo RuvC de Cas9. La inserción de uno o más dominios RuvC o de tipo RuvC en Cas9 puede ser útil para ampliar la especificidad en un ácido nucleico diana. La inserción de uno o más dominios RuvC o de tipo RuvC en Cas9 puede ser útil para duplicar la especificidad en un ácido nucleico diana. Por ejemplo, la inserción de uno o más dominios RuvC o de tipo RuvC puede configurar Cas9 para reconocer un tramo más largo de ácido nucleico diana, reconocer un híbrido de ARN-ADN diferente y/o reconocer un ácido nucleico diana con mayor afinidad de unión.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un dominio RuvC modificado.

Modificaciones de los dominios Cas9 que contienen regiones homólogas de ARN polimerasa

En algunos casos, un polipéptido dirigido a un sitio puede compartir homología con la ARN polimerasa. Ambas proteínas pueden compartir dominios funcionalmente homólogos similares que participe en la catálisis de unión y manipulación de ácidos nucleicos. Por ejemplo, la ARN polimerasa puede comprender regiones de secuencia polipeptídica que participe en la unión de los dúplex de ARN-ADN. En algunos casos, estas regiones pueden ayudar a fundir el dúplex.

En algunos casos, un polipéptido dirigido a un sitio también puede comprender ciertas regiones que afectan a la especificidad de unión de la enzima hacia los ácidos nucleicos. En algunos casos, estas regiones pueden compartir secuencia u homología funcional con dominios o regiones como las encontradas en la ARN polimerasa. En algunos casos, la región básica del extremo N del polipéptido dirigido a un sitio puede unirse al ARNcrtra y al ARNcr o a un solo ARN (ARNgs). En *S. pyogenes*, esto puede corresponder a una región de restos 50-100.

En general, la presente divulgación proporciona cualquier modificación adecuada de esta región o regiones adyacentes. En algunos casos, la región de unión a ARNcrtra/ARNcr (por ejemplo, la región de unión a ácido nucleico dirigida a ácido nucleico) del polipéptido dirigido a un sitio puede modificarse para alterar la especificidad hacia el ácido nucleico. En algunos casos, la región de unión a ARNcrtra/ARNcr puede compartir homología con la región de unión a ARNcrtra/ARNcr identificada en la proteína Cas9 de *S. pyogenes* (restos 50-100). En algunos casos, la región de unión a ARNcrtra/ARNcr puede compartir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 50-100 de la región de unión a ARNcrtra/ARNcr de Cas9 de *S. pyogenes*. En algunos casos, la región de unión a ARNcrtra/ARNcr puede compartir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de homología con los restos 50-100 de la región de unión a ARNcrtra/ARNcr de Cas9 de *S. pyogenes*.

En algunos casos, las modificaciones en la región de unión a ARNcrtra/ARNcr pueden incluir, pero sin limitación, modificaciones de aminoácidos individuales, como se describe en el presente documento. En algunos casos, la modificación de la región de unión a ARNcrtra/ARNcr puede incluir, pero sin limitación, inserciones, eliminaciones o

sustituciones de aminoácidos o polipéptidos individuales, tales como otros elementos proteicos (por ejemplo, dominios, motivos estructurales, proteínas).

Las modificaciones pueden incluir modificaciones de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos de la región de unión a ARNcrtra/ARNcr. Las modificaciones pueden incluir modificaciones, como máximo, en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más aminoácidos de la región de unión a ARNcrtra/ARNcr. Las modificaciones también pueden incluir al menos un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr. Las modificaciones también pueden incluir, como máximo, un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr.

En algunos casos, las modificaciones de las secuencias de la región de unión a ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir determinados motivos estructurales del polipéptido, incluyendo, aunque sin limitación, la hélice alfa, cadena beta, lámina beta, hélice 310, hélice pi, motivo de polipropolina I, motivo de polipropolina II, motivo de poliprolina III, vuelta beta, alfa-vuelta-alfa, o acodamientos o bisagras de hélice.

Por ejemplo, las sustituciones en la región de unión del ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir la sustitución o adición con una o más proteínas o fragmentos de las mismas. Por ejemplo, la región de unión a ARNcrtra/ARNcr podría ser sustituida con el dominio de unión a ARN de cualquiera de los miembros del sistema CRISPR de Tipo I, Tipo II o Tipo III de unión al ARN conocidos. La región de unión a ARNcrtra/ARNcr podría ser sustituida con el dominio de unión a ARN de cualquiera de los miembros de unión al ARN de la superfamilia RAMP. La región de unión a ARNcrtra/ARNcr podría sustituirse con el dominio de unión a ARN de cualquiera de los miembros de unión a ARN conocidos de las familias Cas7, Cas6, Cas5. En un ejemplo, el requisito de ARNcrtra puede reemplazarse con el requisito de una secuencia de horquilla 5' con la secuencia del espaciador colocada cadeba abajo de la horquilla para el reconocimiento de ADN.

En algunos casos, las modificaciones de las secuencias de la región de unión del ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir la eliminación de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr. En algunos casos, las modificaciones de las secuencias de la región de unión a ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir la eliminación de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr.

En algunos casos, las modificaciones de las secuencias de la región de unión del ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de al menos un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr de Cas9 homóloga. En algunos casos, las modificaciones de las secuencias de la región de unión del ARNcrtra/ARNcr del polipéptido dirigido a un sitio pueden incluir la adición o sustitución de, como máximo, un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de la región de unión a ARNcrtra/ARNcr de Cas9 homóloga.

Las regiones de unión a ARNcrtra/ARNcr de polipéptido dirigido a un sitio homólogas pueden derivarse de cualquier organismo adecuado. En algunos casos, la región de unión a ARNcrtra/ARNcr puede derivarse de organismos procariotas, incluyendo, pero sin limitación, arqueas, bacterias, protistas (por ejemplo,, *E. coli, S. pyogenes, S. thermophilus, P.furiosus,* etc.).Por ejemplo, la región de unión a ARNcrtra/ARNcr de Cas9 de *S. pyogenes* puede sustituirse con o modificarse mediante la inserción de la región de unión a ARNcrtra/ARNcr, o un fragmento de la misma, derivada de otra Cas9, tal como una derivada de otra especie.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y un dominio de tipo polimerasa modificado.

Modificaciones para alterar la especificidad hacia PAM

En algunos casos, un polipéptido dirigido a un sitio puede reconocer un motivo adyacente al protoespaciador (PAM). Un PAM puede ser cualquier secuencia de un ácido nucleico diana que es reconocido por un polipéptido dirigido a un sitio y que está inmediatamente a 3' de la secuencia de ácido nucleico diana dirigida por el espaciador de un ácido nucleico dirigido a ácido nucleico. Por ejemplo, un PAM puede comprender 5'-NGG-3' o 5'-NGGNG-3', 5'-NNAAAAW-3', 5'-NNNNGATT-3', 5'-GNNNCNNA-3', 5'-NNNACA-3' donde N es cualquier nucleótido y N está a inmediatamente a 3' de la secuencia de ácido nucleico diana de la secuencia de espaciador.

Un polipéptido dirigido a un sitio puede modificarse para alterar la especificidad hacia PAM. Por ejemplo, un polipéptido dirigido a un sitio puede modificarse de manera que, antes de la modificación, el polipéptido se dirija a un primer motivo adyacente al protoespaciador y, después de la modificación, el polipéptido dirigido a un sitio se dirija a un segundo

motivo adyacente al protoespaciador. En algunos casos, la especificidad hacia PAM alterada puede comprender un cambio en la especificidad de unión (por ejemplo, unión aumentada, unión disminuida) y/o un cambio en la constante de unión (por ejemplo, aumento de K_d, disminución de K_d).

Un polipéptido dirigido a un sitio puede modificarse de manera que el polipéptido dirigido a un sitio pueda reconocer un nuevo tipo de PAM diferente del tipo que reconoce el polipéptido dirigido a un sitio de tipo silvestre. Por ejemplo, un polipéptido dirigido a un sitio que reconoce el PAM 5'-NGG-3' puede modificarse de modo que pueda reconocer el PAM 5'-NGGNG-3', 5'-NNAAAAW-3', 5'-NNNNGATT-3', 5'-GNNNCNNA-3' 5'-NNNACA-3'.

Se puede diseñar cualquier región de un polipéptido dirigido a un sitio (por ejemplo, hélice puente, dominio HNH y/o dominio de tipo HNH, dominio RuvC y/o dominio de tipo RuvC, parche básico) para alterar la especificidad hacia PAM de acuerdo con los métodos de la divulgación.

Las regiones correspondientes a los restos 445-507, 446-497, 1096-1225, 1105-1138 de un polipéptido dirigido a un sitio de tipo silvestre (por ejemplo, Cas9 de *S. pyogenes*, SEQ ID NO: 8) pueden modificarse para modificar el reconocimiento de PAM. El diseño de estas regiones puede comprender mutaciones introductoras, reemplazando con regiones correspondientes de otras Cas9 ortólogas, eliminaciones, inserciones, etc. Las regiones correspondientes a los restos 718-757, 22-49, 65-95, 445-507, 446-497, 1096-1225, 1105-1138 puede diseñarse para modificar el reconocimiento del ácido nucleico dirigido a ácido nucleico. Se pueden diseñar regiones correspondientes a los restos 445-507 y 1105-1138 para modificar el reconocimiento del dominio P.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y una modificación, en donde, antes de la introducción de la modificación, el polipéptido dirigido a un sitio se adapta para unirse a un primer PAM y, después de la introducción de la modificación, el polipéptido dirigido a un sitio se adapta para unirse a un PAM diferente.

Modificaciones para alterar la especificidad del ácido nucleico dirigido a ácido nucleico

En algunos casos, un polipéptido dirigido a un sitio puede reconocer a un ácido nucleico dirigido a ácido nucleico. Un polipéptido dirigido a un sitio puede modificarse para alterar una especificidad hacia el ácido nucleico dirigida a ácido nucleico. Por ejemplo, un polipéptido dirigido a un sitio puede modificarse de manera que, antes de la modificación, el polipéptido se dirija aun primer ácido nucleico dirigido a ácido nucleico y, después de la modificación, el polipéptido dirigido a un sitio se dirija un segundo ácido nucleico dirigido a ácido nucleico. En algunos casos, la especificidad hacia el ácido nucleico dirigido a ácido nucleico alterada puede comprender un cambio en la especificidad de unión (por ejemplo, unión aumentada, unión disminuida) y/o un cambio en la constante de unión (por ejemplo, aumento de K_d, disminución de K_d).

Un polipéptido dirigido a un sitio puede modificarse de manera que el polipéptido dirigido a un sitio pueda reconocer un nuevo tipo de ácido nucleico dirigido a ácido nucleico diferente del tipo que reconoce el polipéptido dirigido a un sitio de tipo silvestre. Se puede diseñar cualquier región de un polipéptido dirigido a un sitio (por ejemplo, hélice puente, dominio HNH y/o dominio de tipo HNH, dominio RuvC y/o dominio de tipo RuvC, parche básico) para alterar la especificidad hacia PAM de acuerdo con los métodos de la divulgación.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y una modificación, en donde, antes de la introducción de la modificación, el polipéptido dirigido a un sitio se adapta para unirse a un primer ácido nucleico dirigido a ácido nucleico y, después de la introducción de la modificación, el polipéptido dirigido a un sitio está adaptado para unirse a un ácido nucleico dirigido a ácido nucleico diferente.

Modificaciones para alterar los requisitos de hibridación

Inserciones

Un polipéptido dirigido a un sitio puede modificarse para aumentar la especificidad de unión a un ácido nucleico diana. Se puede insertar una secuencia en el polipéptido dirigido a un sitio. En algunos casos, se puede insertar un dominio HNH y/o de tipo HNH en un polipéptido dirigido a un sitio. La secuencia no nativa (por ejemplo, dominio HNH y/o de tipo HNH) puede proceder de cualquier especie. La inserción puede tener lugar en cualquier lugar del polipéptido dirigido a un sitio. La inserción puede ocurrir en tándem (por ejemplo, adyacente) al dominio nativo HNH y/o de tipo HNH del polipéptido dirigido a un sitio. El dominio HNH y/o de tipo HNH insertado puede comprender una mutación. El dominio HNH y/o de tipo HNH insertado puede comprender una mutación que reduzca la actividad nucleasa del dominio. En algunos casos, se puede insertar un dominio RuvC y/o de tipo RuvC en un polipéptido dirigido a un sitio. La inserción puede tener lugar en cualquier lugar del polipéptido dirigido a un sitio. La inserción puede ocurrir en tándem (por ejemplo, adyacente) al dominio nativo RuvC y/o de tipo RuvC del polipéptido dirigido a un sitio. El dominio RuvC y/o de tipo RuvC insertado puede comprender una mutación. El dominio RuvC y/o de tipo RuvC insertado puede

comprender una mutación que reduzca la actividad nucleasa del dominio.

Un polipéptido dirigido a un sitio puede modificarse para aumentar la especificidad de unión a un ácido nucleico dirigido a ácido nucleico. Se puede insertar una secuencia en el polipéptido dirigido a un sitio. Se puede insertar un dominio HNH y/o de tipo HNH en un polipéptido dirigido a un sitio. La secuencia no nativa (por ejemplo, dominio HNH y/o de tipo HNH) puede proceder de cualquier especie. La inserción puede tener lugar en cualquier lugar del polipéptido dirigido a un sitio. La inserción puede ocurrir en tándem (por ejemplo, adyacente) al dominio nativo HNH y/o de tipo HNH del polipéptido dirigido a un sitio. El dominio HNH y/o de tipo HNH insertado puede comprender una mutación. El dominio HNH y/o de tipo HNH insertado puede comprender una mutación que reduzca la actividad nucleasa del dominio. Se puede insertar un dominio RuvC y/o de tipo RuvC en un polipéptido dirigido a un sitio. La inserción puede tener lugar en cualquier lugar del polipéptido dirigido a un sitio. La inserción puede ocurrir en tándem (por ejemplo, adyacente) al dominio nativo RuvC y/o de tipo RuvC del polipéptido dirigido a un sitio. El dominio RuvC y/o de tipo RuvC insertado puede comprender una mutación. El dominio RuvC y/o de tipo RuvC insertado puede comprender una mutación que reduzca la actividad nucleasa del dominio.

Un polipéptido dirigido a un sitio puede diseñarse para comprender un dominio de polipéptido que puedA unirse a híbridos de ARN-ADN (por ejemplo, dominio RNasa, dominio de dedo de cinc). Por ejemplo, un polipéptido dirigido a un sitio se puede diseñar para comprender un dominio RNasaH. El dominio RNasaH insertado puede comprender una mutación. El dominio RNasaH insertado puede comprender una mutación que reduzca la actividad nucleasa del dominio.

Se puede diseñar un polipéptido dirigido a un sitio para que comprenda un dominio de polipéptido que pueda unirse a ADN bicatenario (por ejemplo, dominios que comprenden motivos de hélice-vuelta-hélice, dominios que comprenden motivos de cremallera de leucina, dominios que comprenden motivos de hélice-lazo-hélice, dominios que comprenden motivos de dedos de cinc). Por ejemplo, un polipéptido dirigido a un sitio puede diseñarse para comprender un motivo hélice-vuelta-hélice. Los ejemplos no limitantes de motivos de hélice-giro-hélice incluyen los de dnaB, TetR, MuB, P2R, CysB, BirA, el bacteriófago lambda represor, Engrailed, Myb, LuxR, MarR, ETS, ZNF10a, Kox-1. El motivo hélice-lazo-hélice puede ser dihelicoidal, trihelicoidal, tetrahelicoidal, hélice-vuelta-hélice alado u otro hélice-lazo-hélice modificado. El dominio insertado puede comprender una mutación. El dominio insertado puede comprender una mutación que reduzca la actividad nucleasa del dominio.

Mutaciones compensatorias

Un polipéptido dirigido a un sitio puede comprender una mutación y/o estar diseñado de manera que pueda unirse preferentemente a un ácido nucleico dirigido a ácido nucleico mutado y/o diseñado. Dicha mutación de un par de polipéptido dirigido a un sitio y ácido nucleico dirigido a ácido nucleico se puede denominar mutación compensatoria. Por ejemplo, se puede diseñar un polipéptido dirigido a un sitio de manera que su dominio nucleasa (por ejemplo, HNH y/o de tipo HNH, RuvC y/o de tipo RuvC) se reemplaza por un dominio de unión a ácido nucleico (por ejemplo, dominio de unión a ácido nucleico Csy4, Cas5, Cas6). Se puede diseñar un polipéptido dirigido a un sitio de manera que un dominio de unión a ácido nucleico (por ejemplo, dominio de unión al ácido nucleico Csy4, Cas5, Cas6) se inserta en el polipéptido dirigido a un sitio. El polipéptido dirigido a un sitio resultante puede unirse a un ácido nucleico dirigido a ácido nucleico que está mutado y/o diseñado para comprender un sitio de unión al dominio de unión a ácido nucleico (por ejemplo, sitio de unión para los dominios de unión a ácido nucleico Csy4, Cas5, Cas6). El ácido nucleico dirigido a ácido nucleico puede mutarse y/o diseñarse para que comprenda un sitio de unión al dominio de unión a ácido nucleico en la secuencia mínima de ARNcrtra. El ácido nucleico dirigido a ácido nucleico puede mutarse y/o diseñarse para que comprenda un sitio de unión al dominio de unión a ácido nucleico en la secuencia de ARNcrtra de 3'. El ácido nucleico dirigido a ácido nucleico puede mutarse y/o diseñarse para que comprenda un sitio de unión al dominio de unión a ácido nucleico en la extensión de ARNcrtra.

En algunos casos, un polipéptido dirigido a un sitio comprende una secuencia de aminoácidos que comprende al menos un 15 % de identidad de aminoácidos con un Cas9 de *S. pyogenes*, dos dominios de escisión de ácido nucleico (es decir, un dominio HNH y un dominio RuvC) y una mutación compensatoria, en donde el polipéptido dirigido a un sitio es tal que puede unirse a un ácido nucleico dirigido a ácido nucleico diseñado, pero no a un ácido nucleico dirigido a ácido nucleico no modificado.

Métodos de generación de extremos adhesivos y cortes romos

En algunos casos, se pueden usar una o más nickasas (es decir, un polipéptido dirigido a un sitio que comprende un dominio nucleasa sustancialmente inactivo) para generar cortes bicatenarios dirigidos en ácido nucleico diana. Cada nickasa de una o más nickasas puede dirigirse a una cadena del ácido nucleico diana bicatenario. En algunos casos, se pueden usar dos nickasas para generar un corte bicatenario diana.

Las dos nickasas pueden cortar el ácido nucleico diana generando un corte de extremo romo (en donde los sitios de corte del ácido nucleico diana están en la misma ubicación en cada cadena). Las dos nickasas pueden cortar el ácido nucleico diana en diferentes ubicaciones dentro de cada cadena de manera que permanezcan algunos nucleótidos monocatenarios, generando así un extremo adhesivo.

La escisión del ácido nucleico diana por dos polipéptidos modificados dirigidos a un sitio que tienen actividad de nickasa puede usarse para incurrir en eliminaciones o inserciones de material de ácido nucleico de un ácido nucleico diana escindiendo el ácido nucleico diana y permitiendo que la célula repare la secuencia en ausencia de un polinucleótido donante proporcionado exógenamente. En algunos casos, los métodos de la divulgación pueden usarse para desactivar un gen. Si un ácido nucleico dirigido a ácido nucleico y dos polipéptidos modificados dirigidos a un sitio que tienen actividad de nickasa se administran conjuntamente a las células con una secuencia de polinucleótido donante que incluye al menos un segmento con homología con el ácido nucleico diana, se puede obtener un nuevo material de ácido nucleico insertado/copiado en el sitio. Dichos métodos se pueden usar para agregar, es decir, insertar o reemplazar, material de ácido nucleico a un ácido nucleico diana (por ejemplo, "insertar" un ácido nucleico que codifica una proteína, un ARNip, un miARN, etc.), para agregar un marcador (por ejemplo, 6xHis, una proteína fluorescente (por ejemplo, una proteína fluorescente verde; una proteína fluorescente amarilla, etc.), hemaglutinina (HA), FLAG, etc.), para agregar una secuencia reguladora a un gen (por ejemplo, promotor, señal de poliadenilación, secuencia interna de entrada al ribosoma (IRES), péptido 2A, codón de inicio, codón de terminación, señal de corte y empalme, señal de localización, etc.), para modificar una secuencia de ácido nucleico (por ejemplo, introducir una mutación) y similares.

La **Figura 32** muestra un método para la generación extremos romos por nickasas. El dúplex **3210** de ácido nucleico diana puede comprender una pluralidad de secuencias **3215** de PAM (recuadro), en donde un PAM está en una cadena del ácido nucleico diana **3210** y un PAM está en la otra cadena del ácido nucleico diana **3210**. Un ácido nucleico dirigido a ácido nucleico **3205** como parte de un complejo con la nickasa (no se muestra la nickasa) puede hibridarse con la secuencia del espaciador adyacente al PAM **3215** en cada cadena del ácido nucleico diana **3210**. La nickasa puede escindir una cadena del ácido nucleico diana **3210**. La escisión se indica por los triángulos. Si los PAM están espaciados adecuadamente, las nickasas pueden cortar el ácido nucleico diana en sustancialmente el mismo lugar de cada cadena, lo que da lugar a un extremo romo. Las secuencias de PAM pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 o 50, o más, nucleótidos. Las secuencias de PAM pueden estar separadas, como máximo, por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 o 50, o más, nucleótidos. En algunos casos, los PAM están separados por 6 nucleótidos (es decir, hay 6 nucleótidos entre cada PAM). En algunos casos, el ácido nucleico dirigido a ácido nucleico escinde aproximadamente 3 nucleótidos de 5’ del PAM.

En algunas realizaciones, se pueden emplear dos o más nickasas para generar extremos adhesivos. La **Figura 33** ilustra cómo dos nickasas dirigidas a regiones superpuestas en un ácido nucleico diana pueden producir una rotura bicatenaria escalonada que da lugar a extremos adhesivos. El dúplex **3310** de ácido nucleico diana puede comprender una pluralidad de secuencias **3315** de PAM (recuadro). Un ácido nucleico dirigido a ácido nucleico **3305** como parte de un complejo con la nickasa (no se muestra la nickasa) puede hibridarse con la secuencia del espaciador adyacente al PAM **3315** en cada cadena del ácido nucleico diana **3310**. La nickasa puede escindir una cadena del ácido nucleico diana **3310**. La escisión se indica por los triángulos. Si los PAM están espaciados adecuadamente, las nickasas pueden cortar el ácido nucleico diana en ubicaciones escalonadas, lo que da lugar a un extremo adhesivo. Las secuencias de PAM pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las secuencias de PAM pueden estar separadas, como máximo, por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. La distancia de las secuencias de PAM puede estar relacionada con la longitud del extremo adhesivo generado. Por ejemplo, cuanto más alejados estén los PAM entre sí, más largo será el extremo adhesivo.

Un método para generar extremos adhesivos usando dos o más nickasas puede implicar secuencias de PAM sustancialmente adyacentes entre sí (aunque en cadenas opuestas). En algunos casos, las secuencias de PAM están separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. En algunos casos, las secuencias de PAM están, como máximo, por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. En algunos casos, las secuencias de PAM están separadas por un nucleótido. En algunos casos, las secuencias de PAM no están separadas por ningún nucleótido.

Métodos de enriquecimiento y secuenciación de ácidos nucleicos diana

Visión general

La secuenciación puede ser útil para diagnosticar enfermedades mediante la identificación de mutaciones y/u otras variantes de secuencia (por ejemplo, polimorfismos). Los métodos de la divulgación proporcionan métodos, kits y composiciones para enriquecer una secuencia de ácido nucleico diana sin el uso de metodologías de amplificación. Un ácido nucleico diana puede ser enriquecido con el uso de un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico.

La **FIGURA 3** representa un método ilustrativo de la divulgación. Un polipéptido dirigido a un sitio **305** puede unirse a un ácido nucleico dirigido a ácido nucleico **310**, formando así un complejo **306**. El ácido nucleico diana de ácido nucleico 310 puede comprender un marcador de afinidad hacia ácido nucleico **311**. El polipéptido dirigido a un sitio **305** puede comprender un dominio nucleasa. El polipéptido **305** dirigido a un sitio puede ser enzimáticamente activo. El polipéptido dirigido a un sitio **305** puede comprender un marcador de afinidad **315**. El ácido nucleico dirigido a ácido nucleico **310** puede hibridarse con un ácido nucleico diana **320**. En algunos aspectos, una pluralidad de complejos

306 puede hibridarse con una pluralidad de ubicaciones dentro de un ácido nucleico diana **320**. En una etapa de escisión **325**, el dominio nucleasa de un polipéptido dirigido a un sitio **305** puede escindir o cortar **330** el ácido nucleico diana **320**. El ácido nucleico diana cortado **340** puede purificarse en una etapa de purificación **335**. Los adaptadores **345** pueden ligarse al ácido nucleico diana escindido. Los adaptadores pueden facilitar la secuenciación del ácido nucleico diana escindido.

La **FIGURA 4** representa un método ilustrativo de la divulgación. Un polipéptido dirigido a un sitio **405** puede interactuar con un ácido nucleico dirigido a ácido nucleico **410**, formando así un complejo **406**. El polipéptido dirigido a un sitio **405** puede comprender un dominio nucleasa. En algunos aspectos, el dominio nucleasa del polipéptido dirigido a un sitio **405** puede ser enzimáticamente inactivo. El polipéptido dirigido a un sitio **405** puede comprender un marcador de afinidad **415**. El ácido nucleico dirigido a ácido nucleico **410** puede hibridarse con un ácido nucleico diana **420**. El ácido nucleico diana de ácido nucleico **410** puede comprender un marcador de afinidad hacia ácido nucleico **411**. El marcador de afinidad **411** del ácido nucleico dirigido a ácido nucleico puede comprender una estructura de horquilla. Una pluralidad de complejos **406** puede hibridarse con una pluralidad de ubicaciones dentro de un ácido nucleico diana **420**. En una etapa de fragmentación **225**, el ácido nucleico diana **420** puede fragmentarse en el fragmento de ácido nucleico diana **445** (también denominado en el presente documento "ácido nucleico diana"). El polipéptido dirigido a un sitio **405** puede purificarse mediante un agente de captura **440** que puede unirse al marcador de afinidad **415** del polipéptido dirigido a un sitio **405**. El ácido nucleico diana fragmentado **445** puede eluirse del complejo **406** en una etapa de purificación **450**. En la misma etapa, u opcionalmente, en un etapa diferente, los adaptadores **455** pueden ligarse al ácido nucleico diana escindido. Los adaptadores pueden facilitar la secuenciación del ácido nucleico diana.

Complejo de un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio

Un ácido nucleico dirigido a ácido nucleico puede interactuar con un polipéptido dirigido a un sitio (por ejemplo, una nucleasa guiada a ácido nucleico, por ejemplo, Cas9), formando así un complejo. El ácido nucleico dirigido a ácido nucleico puede guiar al polipéptido dirigido a un sitio a un ácido nucleico diana.

En algunos aspectos, un ácido nucleico dirigido a ácido nucleico puede modificarse de forma tal que el complejo (por ejemplo, que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico) pueda unirse fuera del sitio de escisión del polipéptido dirigido a un sitio. En este caso, el ácido nucleico diana puede no interactuar con el complejo, y el ácido nucleico diana se puede escindir (por ejemplo, libre del complejo).

En algunos aspectos, un ácido nucleico dirigido a ácido nucleico puede diseñado de modo que el complejo pueda unirse dentro del sitio de escisión del polipéptido dirigido a un sitio. En este caso, el ácido nucleico diana puede interactuar con el complejo y el ácido nucleico diana puede unirse (por ejemplo, unirse al complejo).

El ácido nucleico dirigido a ácido nucleico se puede diseñar de manera que el complejo (por ejemplo, que comprende un polipéptido dirigido a un sitio y/o un ácido nucleico dirigido a ácido nucleico) pueda hibridarse en una pluralidad de ubicaciones dentro de una muestra de ácido nucleico.

Se puede poner en contacto una pluralidad de complejos con una muestra de ácido nucleico. La pluralidad de complejos puede comprender ácidos nucleicos dirigidos a ácido nucleico diseñados para hibridarse con la misma secuencia. La pluralidad de complejos puede comprender ácidos nucleicos dirigidos a ácido nucleico diseñados para hibridarse con las diferentes secuencias.

Las secuencias pueden estar en diferentes ubicaciones dentro de un ácido nucleico diana. Las ubicaciones pueden comprender las mismas secuencias de ácido nucleico diana, o similares. Las ubicaciones pueden comprender diferentes secuencias de ácido nucleico diana. Las ubicaciones pueden estar a una distancia definida entre sí. Las ubicaciones pueden estar separadas por menos de 10 kilobases (Kb), menos de 8 Kb, menos de 6 Kb, menos de 4 Kb, menos de 2 Kb, menos de 1 Kb, menos de 900 nucleótidos, menos de 800 nucleótidos, menos de 700 nucleótidos, menos de 600 nucleótidos, menos de 500 nucleótidos, menos de 400 nucleótidos, menos de 300 nucleótidos, menos de 200 nucleótidos, menos de 100 nucleótidos.

Los complejos pueden escindir el ácido nucleico diana que puede producir un ácido nucleico diana escindido que puede tener menos de 10 kilobases (Kb) de longitud, menos de 8 Kb de longitud, menos de 6 Kb de longitud, menos de 4 Kb de longitud, menos de 2 Kb de longitud, menos de 1 Kb de longitud, menos de 900 nucleótidos de longitud, menos de 800 nucleótidos de longitud, menos de 700 nucleótidos de longitud, menos de 600 nucleótidos de longitud, menos de 500 nucleótidos de longitud, menos de 400 nucleótidos de longitud, menos de 300 nucleótidos de longitud, menos de 200 nucleótidos de longitud, menos de 100 nucleótidos de longitud.

Los complejos se pueden unir a un ácido nucleico diana fragmentado que puede tener menos de 10 kilobases (Kb) de longitud, menos de 8 Kb de longitud, menos de 6 Kb de longitud, menos de 4 Kb de longitud, menos de 2 Kb de longitud, menos de 1 Kb de longitud, menos de 900 nucleótidos de longitud, menos de 800 nucleótidos de longitud, menos de 700 nucleótidos de longitud, menos de 600 nucleótidos de longitud, menos de 500 nucleótidos de longitud, menos de 400 nucleótidos de longitud, menos de 300 nucleótidos de longitud, menos de 200 nucleótidos de longitud, menos de 100 nucleótidos de longitud.

Métodos de detección de sitios de unión fuera de la diana de polipéptidos dirigidos a un sitio

Visión general

La presente divulgación describe métodos, composiciones, sistemas y/o kits de determinación de sitios de unión fuera de la diana de polipéptidos dirigidos a un sitio. En algunos aspectos de la divulgación, un polipéptido dirigido a un sitio puede comprender un ácido nucleico dirigido a ácido nucleico, formando así un complejo. El complejo puede ponerse en contacto con un ácido nucleico diana. El ácido nucleico diana puede capturarse con agentes de captura que pueden unirse a los marcadores de afinidad del complejo. La identidad del ácido nucleico diana se puede determinar mediante secuenciación. La secuenciación (por ejemplo, secuenciación de alto rendimiento, por ejemplo, Illumina, Ion Torrent) también puede identificar la frecuencia de sitios de unión fuera de la diana del polipéptido dirigido a un sitio y/o complejo, contando el número de veces que se lee un determinado sitio de unión. Los métodos, las composiciones, los sistemas y/o los kits de la divulgación pueden facilitar el desarrollo de polipéptidos dirigidos a un sitio de manera más precisa y específica.

La **FIGURA 5** representa un método ilustrativo de la divulgación. Un polipéptido dirigido a un sitio **505** puede comprender un marcador de afinidad **510**. El polipéptido dirigido a un sitio puede comprender un dominio de unión a ácido nucleico **515**. El dominio de unión a ácido nucleico **515** puede ser un ácido nucleico. En algunos aspectos, el dominio de unión a ácido nucleico **515** y el polipéptido **505** dirigido a un sitio forman un complejo **531**. El complejo **531** puede ponerse en contacto **525** con un ácido nucleico diana **530**. En un aspecto preferido, el ácido nucleico diana **530** es ADN (por ejemplo, ADN genómico o ADNg). El complejo puede purificarse por afinidad **535** con un agente de captura **540**. El agente de captura **540** puede unirse al marcador de afinidad **510** del polipéptido dirigido a un sitio **505**. El agente de captura **540** puede comprender un segundO marcador de afinidad **545**. El agente de captura **540** puede ser purificado por afinidad **550** uniéndose a un soporte sólido **555**. En algunos aspectos, el soporte sólido **555** es una perla recubierta con un reactivo de afinidad que puede unirse al marcador de afinidad **545** del agente de captura. Opcionalmente, el soporte sólido **555** puede unirse al marcador de afinidad **510** del polipéptido dirigido a un sitio **505** para facilitar la purificación. En algunos aspectos, se pueden producir una o más series de purificación. Cada serie puede comprender poner en contacto un soporte sólido **555** con los marcadores de afinidad del polipéptido dirigido a un sitio **510** y/o el agente de captura **545**. El complejo purificado por afinidad puede eluirse del ácido nucleico diana **530**. El ácido nucleico diana puede prepararse posteriormente para su posterior procesamiento. El procesamiento puede incluir métodos de análisis cadena abajo, por ejemplo, secuenciación.

La **FIGURA 6** representa un método ilustrativo de la divulgación. Un polipéptido dirigido a un sitio **605** puede comprender un marcador de afinidad **610**. El polipéptido dirigido a un sitio **605** puede comprender un dominio de unión a ácido nucleico **615**. El dominio de unión a ácido nucleico **615** puede ser un ácido nucleico. En algunos aspectos, el dominio de unión a ácido nucleico **615** puede comprender un marcador de afinidad **620**. En algunos aspectos, el dominio de unión a ácido nucleico **615** y el polipéptido específico del sitio **605** pueden formar un complejo **631**. El complejo **631** puede ponerse en contacto **625** con un ácido nucleico diana **630**. En un aspecto preferido, el ácido nucleico diana **630** es ADN. El complejo **631** puede purificarse por afinidad **635** con un agente de captura **640**. El agente de captura **640** puede ser un polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo. El agente de captura **640** puede ser una variante condicionalmente enzimáticamente inactiva de Csy4. El agente de captura **640** puede unirse al marcador de afinidad **620**. El agente de captura **640** puede comprender un marcador de afinidad **645**. El agente de captura **640** puede ser purificado por afinidad **650** uniéndose a un soporte sólido **655**. En algunos aspectos, el soporte sólido es una perla recubierta con un reactivo de afinidad que puede unirse al marcador de afinidad **645** del agente de captura **640**. Opcionalmente, el soporte sólido **655** puede unirse al marcador de afinidad **610** del polipéptido dirigido a un sitio **605** para facilitar la purificación. En algunos aspectos, pueden producirse dos series de purificación, comprendiendo cada serie poner en contacto un soporte sólido **655** con los marcadores de afinidad del polipéptido dirigido a un sitio **610** y/o el agente de captura **640**. La escisión del marcador de afinidad 620 puede facilitar la elución **660** del ácido nucleico diana **630** del soporte sólido **655**. El ácido nucleico diana **630** puede prepararse posteriormente para métodos de análisis posteriores cadena abajo tales como secuenciación.

Métodos

La divulgación proporciona métodos para la inmunoprecipitación y secuenciación de nucleasas (NIP-Seq). En algunos aspectos, el método puede comprender a) poner en contacto una muestra de ácido nucleico con un complejo que comprende un polipéptido dirigido a un sitio enzimáticamente inactivo, un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico. El complejo puede hibridarse con un ácido nucleico diana. El complejo puede capturarse con un agente de captura, y el ácido nucleico diana unido al complejo puede secuenciarse. En algunos aspectos, el método puede comprender además determinar la identidad del sitio de unión fuera de la diana. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

Los métodos se pueden realizar fuera de una célula. Por ejemplo, una muestra puede comprender ADN genómico purificado, lisado de células, tejido homogeneizado, plasma y similares. Los métodos se pueden realizar en células.

Los complejos de polipéptido dirigido a un sitio-ácido nucleico dirigido a ácido nucleico se pueden fijar o entrecruzar para formar complejos. Las células se pueden entrecruzar antes del lisado. Las células fijas o entrecruzadas pueden estabilizar los complejos de proteína-ADN en la célula. Los fijadores y agentes de entrecruzamiento adecuados pueden incluir, formaldehído, glutaraldehído, fijadores a base de etanol, fijadores a base de metanol, acetona, ácido acético, tetraóxido de osmio, dicromato de potasio, ácido crómico, permanganato de potasio, mercuriales, picratos, formol, paraformaldehído, agentes de entrecruzamiento de NHS-éster reactivos con amina tales como bis[sulfosuccinimidil]suberato (BS3), 3,3'-ditiobis[sulfosuccinimidilpropionato] (DTSSP), bis[sulfosuccinimidilsuccinato de etilenglicol (sulfo-EGS), glutarato de disuccinimidilo (DSG), propionato de ditiobis[succinimidilo] (DSP), suberato de disuccinimidilo (DSS), bis[succinimidilsuccinato] de etilenglicol (EGS), agentes de entrecruzamiento de NHS-éster/diazirina tales como NHS-diazirina, NHS-LC-diazirina, NHS-SS-diazirina, sulfo-NHS-diazirina, sulfo-NHS-LC-diazirina y sulfo-NHS-SS-diazirina.

El ácido nucleico (por ejemplo, ADN genómico) puede tratarse para fragmentar el ADN antes de la purificación por afinidad. La fragmentación se puede realizar a través de métodos físicos, mecánicos o enzimáticos. La fragmentación física puede incluir la exposición de un polinucleótido diana al calor o a la luz ultravioleta (UV). La disrupción mecánica puede usarse para cortar mecánicamente un polinucleótido diana en fragmentos del intervalo deseado. La cizalla mecánica se puede lograr a través de varios métodos conocidos en la técnica, que incluyen pipeteo repetitivo del polinucleótido diana, sonicación y nebulización. Los polinucleótidos diana también pueden fragmentarse usando métodos enzimáticos. En algunos casos, se puede realizar la digestión enzimática usando enzimas tales como las enzimas de restricción. Las enzimas de restricción pueden usarse para realizar una fragmentación específica o no específica de polinucleótidos diana. Los métodos pueden usar uno o más tipos de enzimas de restricción, descritas en general como enzimas de Tipo I, enzimas de Tipo II y/o enzimas de Tipo III. Las enzimas de Tipo II y Tipo III, en general, están disponibles en el mercado, y son bien conocidas en la técnica. Las enzimas de Tipo II y de Tipo III reconocen secuencias específicas de nucleótidos nucleotídicos dentro de una secuencia polinucleotídica bicatenaria (una "secuencia de reconocimiento" o "sitio de reconocimiento"). Tras la unión y el reconocimiento de estas secuencias, las enzimas de Tipo II y Tipo III escinden la secuencia de polinucleótidos. En algunos casos, la escisión dará lugar a un fragmento de polinucleótido con una parte de ADN monocatenario saliente, denominada "extremo adhesivo". En otros casos, la escisión no dará lugar a un fragmento con un saliente, creando un "extremo romo". Los métodos pueden comprender el uso de enzimas de restricción que generen extremos adhesivos o extremos romos. Los fragmentos de ácidos nucleicos también se pueden generar mediante técnicas de amplificación (por ejemplo, reacción en cadena de la polimerasa, reacción en cadena de la polimerasa de largo alcance, reacción en cadena de la polimerasa lineal, etc.).

Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio pueden purificarse mediante incubación con un soporte sólido. Por ejemplo, si el polipéptido dirigido a un sitio comprende un marcador de biotina, el soporte sólido puede recubrirse con avidina o estreptavidina para unirse al marcador de biotina.

En algunos aspectos, Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, el ácido nucleico diana y/o el ácido nucleico dirigido a ácido nucleico se purifican por incubación con un agente de captura. Un agente de captura puede referirse a cualquier agente que pueda unirse a un marcador de afinidad fusionado con el polipéptido dirigido a un sitio. Los ejemplos de agentes de captura pueden incluir biotina, estreptavidina y anticuerpos. Por ejemplo, si el marcador de afinidad fusionado con el polipéptido dirigido a un sitio es un marcador FLAG, entonces el agente de captura será un anticuerpo anti-marcador FLAG. En algunos aspectos, el agente de captura puede comprender un marcador de afinidad (por ejemplo, biotina,estreptavidina).

En algunos casos, el agente de captura es una endorribonucleasa enzimáticamente inactiva. Por ejemplo, un agente de captura puede ser un polipéptido dirigido a un sitio enzimáticamente inactivo, una Csy4, Cas5 o Cas6 enzimáticamente inactiva.

El agente de captura se puede purificar con un soporte sólido. Por ejemplo, si el agente de captura comprende un marcador de biotina, la perla puede recubrirse con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos del método, se pueden realizar dos o más series de purificación. Se pueden realizar al menos 1, 2, 3, 4, 5, 6, 7 o más series de purificación. Se pueden realizar, como máximo, 1, 2, 3, 4, 5, 6, 7 o más series de purificación. Una primera serie de purificación puede comprender la purificación con un soporte sólido que pueda unirse al marcador de afinidad del agente de captura, y una segunda serie de purificación puede comprender la purificación con un soporte sólido que pueda unirse al marcador de afinidad del polipéptido dirigido a un sitio. Una primera serie de purificación puede comprender la purificación con un soporte sólido que pueda unirse al marcador de afinidad del polipéptido dirigido a un sitio, y una segunda serie de purificación puede comprender la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura. El método puede usarse para optimizar la especificidad de unión de un polipéptido dirigido a un sitio realizando el método más de una vez.

El complejo capturado puede comprender un polipéptido dirigido a un sitio y un ácido nucleico diana. El ácido nucleico diana se puede eluir del complejo del polipéptido dirigido a un sitio mediante métodos tales como lavado con alto contenido de sal, precipitación en etanol, ebullición y purificación en gel.

El ADN eluido puede prepararse para el análisis de secuenciación (por ejemplo, cizalla, ligadura de adaptadores). La preparación para el análisis de secuenciación puede incluir la generación de bancos de secuenciación del ácido nucleico diana eluido. El análisis de secuenciación puede determinar la identidad y frecuencia de los sitios de unión fuera de la diana de los polipéptidos dirigidos a un sitio. La determinación de las secuencias también se realizará usando métodos que determinen muchas secuencias de ácido nucleico (normalmente, de miles a miles de millones) de una manera intrínsecamente paralela, donde se leen muchas secuencias preferentemente en paralelo mediante un proceso en serie de alto rendimiento. Dichos métodos incluyen, pero sin limitación, pirosecuenciación (por ejemplo, como la comercializada por 454 Life Sciences, Inc., Branford, Conn.); secuenciación por ligadura (por ejemplo, como la comercializada en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, Calif.); La secuenciación por síntesis usando nucleótidos modificados (tales como los comercializados en la tecnología TruSeq™ y HiSeq™ por Illumina, Inc., San Diego, Calif., HeliScope™ de Helicos Biosciences Corporation, Cambridge, Mass. y PacBio RS de Pacific Biosciences of California, Inc., Menlo Park, Calif.), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, Calif.); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, Calif.); tecnologías de secuenciación basadas en nanoporos (por ejemplo, las desarrolladas por Oxford Nanopore Technologies, LTD, Oxford, Reino Unido) y otros métodos de secuenciación altamente paralelizados conocidos.

En algunos aspectos, el método comprende además recopilar datos y almacenar datos. Los datos pueden ser legibles por máquina y pueden almacenarse y/o recopilarse en el servidor de un ordenador (por ejemplo, **Figura 31** y Ejemplo 27).

Métodos de detección de variantes de secuencia en ácidos nucleicos

Visión general

En algunos aspectos, los métodos de la divulgación proporcionan la detección de variantes de secuencia en ácidos nucleicos. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. Como se representa en la **Figura 7**, una muestra de ácido nucleico **705** se puede ligar **720** con un marcador de ácido nucleico **710**. El marcador de ácido nucleico puede ser un ARN guía sencillo. El marcador de ácido nucleico puede comprender un ARNcr. El marcador de ácido nucleico puede comprender un marcador detectable **715**. En conjunto, la muestra de ácido nucleico **705** ligada al marcador de ácido nucleico **710** puede denominarse muestra de prueba marcada **721**. La muestra de prueba marcada **721** puede ponerse en contacto **725** con una matriz **740** que comprenda oligonucleótidos inmovilizados **735**. Los oligonucleótidos inmovilizados **735** pueden denominarse banco de ácido nucleico. Los oligonucleótidos **735** pueden ser ADN bicatenario. Los oligonucleótidos **735** pueden comprender un marcador detectable **730**. Los miembros individuales de la muestra de prueba marcada **721** pueden hibridarse **745** con los oligonucleótidos **735** con los que comparten suficiente complementariedad para facilitar la hibridación. La cantidad de hibridación puede cuantificarse comparando las intensidades de los dos marcadores detectables **715** y **730**. Por ejemplo, los oligonucleótidos hibridados pueden mostrar dos marcadores detectables. Los oligonucleótidos no hibridados pueden mostrar un marcador detectable **730**. La muestra hibridada puede ponerse en contacto con un polipéptido dirigido a un sitio **750**. El polipéptido dirigido a un sitio puede escindir **755** los oligonucleótidos **735** en la matriz **740** que se han hibridado con miembros de la muestra de prueba marcada **721**. La escisión por el polipéptido dirigido a un sitio puede permitir que los miembros hibridados de la muestra de prueba marcada **721** sean eliminados. Tras la escisión por el polipéptido dirigido a un sitio **750**, solo permanecerán en la matriz los marcadores detectables **760** de oligonucleótidos no hibridados. El marcador detectable **760** restante puede cuantificarse. La cuantificación de los marcadores detectables **760** restantes se puede correlacionar con qué secuencias se representaron en la muestra de ácido nucleico **705** y cuáles no. Los oligonucleótidos que no muestran un marcador detectable **760** restante corresponden a secuencias representadas en la muestra de ácido nucleico **705**. Los oligonucleótidos que muestran un marcador detectable **760** restante corresponden a secuencias que no estaban representadas en la muestra de ácido nucleico **705**.

En algunos aspectos, una muestra de ácido nucleico **805** se puede ligar **820** con un marcador de ácido nucleico **810**. El marcador de ácido nucleico puede ser un ARN guía sencillo. El marcador de ácido nucleico puede comprender un ARNcr. El marcador de ácido nucleico puede comprender un marcador detectable **815**. En conjunto, la muestra de ácido nucleico ligada al marcador de ácido nucleico puede denominarse muestra de prueba marcada **821**. La muestra de prueba marcada **821** puede ponerse en contacto **825** con una matriz **840** que comprenda oligonucleótidos inmovilizados **835**. Los oligonucleótidos inmovilizados pueden denominarse banco de ácido nucleico. Los oligonucleótidos **835** pueden ser ADN bicatenario. Los miembros individuales de la muestra de prueba marcada **821** pueden hibridarse **845** con los oligonucleótidos **835** con los que comparten suficiente complementariedad para facilitar la hibridación. La muestra hibridada puede ponerse en contacto con un polipéptido dirigido a un sitio **850**. El polipéptido dirigido a un sitio puede escindir **855** los oligonucleótidos **835** en la matriz **840** que se han hibridado con miembros de la muestra de prueba marcada **821**. La escisión por el polipéptido dirigido **850** a un sitio puede permitir que los miembros hibridados de la muestra de prueba marcada **821** sean eliminados. La escisión por el polipéptido dirigido a un sitio **850** puede permitir que una parte del oligonucleótido inmovilizado sea escindida y separada de la matriz **860**. Los oligonucleótidos escindidos separados **860** pueden ligarse **865** a los adaptadores apropiados **870** para la

secuenciación. La secuenciación de los oligonucleótidos escindidos **860** puede determinar las secuencias representadas en la muestra de ácido nucleico **805**.

En algunos aspectos, se puede generar un banco de ácido nucleico para el análisis de secuenciación usando plataformas de secuenciación de alto rendimiento disponibles en el mercado. El banco puede comprender ácidos nucleicos que pueden comprender una o más marcadores de secuenciación **930** y una secuencia diana **945**. La secuencia diana **945** puede ser una secuencia que se puede representar en una muestra de ácido nucleico **905**. Una secuencia diana **945** puede comprender una secuencia de motivo adyacente al protoespaciador (PAM). Opcionalmente, los ácidos nucleicos de un banco de ácido nucleico pueden comprender una o más secuencias de polinucleótidos de identificación **935**, y una o más secuencias de extensión **940**. En este aspecto, una muestra de ácido nucleico **905** se puede ligar **920** con un marcador de ácido nucleico **910**. El marcador de ácido nucleico puede ser un ARN guía sencillo. El marcador de ácido nucleico puede comprender un ARNcr. Opcionalmente, el marcador de ácido nucleico puede comprender un marcador de afinidad **915**. En conjunto, la muestra de ácido nucleico ligada al marcador de ácido nucleico puede denominarse muestra de prueba marcada **921**. La muestra de prueba marcada **921** puede ponerse en contacto **925** con un banco de ácido nucleico. La muestra de prueba marcada **921** puede hibridarse con un ácido nucleico en el banco de ácido nucleico, formando un complejo **946**. La muestra de prueba marcada hibridada y el banco de ácido nucleico pueden ponerse en contacto con un polipéptido dirigido a un sitio **950**. El polipéptido dirigido a un sitio **950** puede escindir los miembros del banco de ácido nucleico hibridado. Los miembros del banco de ácido nucleico escindidos **965** pueden separarse de los miembros no escindidos. Los miembros no escindidos pueden ser sometidos a análisis de secuencia. El análisis de secuenciación puede determinar qué secuencias se representaron en la muestra de ácido nucleico **905**. Por ejemplo, las secuencias de los miembros no escindidos pueden corresponder a secuencias que no se representaron en la muestra de ácido nucleico **905**. Estas secuencias se pueden eliminar de las secuencias conocidas del banco de ácido nucleico. Las secuencias resultantes pueden ser las secuencias de los miembros escindidos **965** del banco de ácido nucleico que pueden corresponder a secuencias que se representaron en la muestra de ácido nucleico **905**.

El polipéptido dirigido a un sitio **950** puede comprender un marcador de afinidad **955**. Opcionalmente, el polipéptido dirigido a un sitio **950** puede ser una variante enzimáticamente inactiva de un polipéptido dirigido a un sitio. En algunos aspectos, un polipéptido dirigido a un sitio enzimáticamente inactivo puede ponerse en contacto con un banco de ácido nucleico hibridado (por ejemplo, complejo **946**). El polipéptido dirigido a un sitio puede unirse pero no puede escindir los miembros del banco de ácido nucleico hibridado. El polipéptido dirigido a un sitio puede purificarse por afinidad **970** con un agente de captura **975** que puede unirse al marcador de afinidad **955**. Opcionalmente, el complejo **946** puede purificarse por afinidad con un agente de captura que pueda unirse al marcador de afinidad **915**. Los miembros del banco de ácido nucleico purificado por afinidad pueden someterse a análisis de secuenciación. En este aspecto, los miembros del banco de ácido nucleico secuenciados pueden corresponder a secuencias que están representadas en la muestra de ácido nucleico **905**.

Secuenciación

Los métodos de detección de variantes de secuencia pueden comprender la secuenciación de las variantes. La determinación de las secuencias también se puede realizar usando métodos que determinen muchas secuencias de ácido nucleico (normalmente, de miles a miles de millones) de una manera intrínsecamente paralela, donde se leen muchas secuencias preferentemente en paralelo mediante un proceso en serie de alto rendimiento. Dichos métodos pueden incluir, pero sin limitación, pirosecuenciación (por ejemplo, como la comercializada por 454 Life Sciences, Inc., Branford, Conn.); secuenciación por ligadura (por ejemplo, como la comercializada en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, Calif.); La secuenciación por síntesis usando nucleótidos modificados (tales como los comercializados en la tecnología TruSeq™ y HiSeq™ por Illumina, Inc., San Diego, Calif., HeliScope™ de Helicos Biosciences Corporation, Cambridge, Mass. y PacBio RS de Pacific Biosciences of California, Inc., Menlo Park, Calif.), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, Calif.); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, Calif.); tecnologías de secuenciación basadas en nanoporos (por ejemplo, las desarrolladas por Oxford Nanopore Technologies, LTD, Oxford, Reino Unido), secuenciación capilar (por ejemplo, comercializada en MegaBACE por Molecular Dynamics), secuenciación electrónica, secuenciación de moléculas individuales (por ejemplo, tal como la comercializada en tecnología SMRT™ de Pacific Biosciences, Menlo Park, Calif.), secuenciación microfluídica de gotas, secuenciación por hibridación (tal como la comercializada por Affymetrix, Santa Clara, California), secuenciación de bisulfato y otros métodos de secuenciación altamente paralelizados conocidos.

PCR en tiempo real

Los métodos de detección de variantes de secuencia pueden comprender la detección de las variantes usando PCR en tiempo real. La determinación de la secuencia se puede realizar mediante una reacción en cadena de la polimerasa en tiempo real (RT-PCR, también conocida como PCR cuantitativa (QPCR)) que puede detectar una cantidad de ácido nucleico amplificable presente en una muestra. La QPCR es una técnica basada en la reacción en cadena de la polimerasa, y puede usarse para amplificar y cuantificar simultáneamente un ácido nucleico diana. La QPCR puede permitir tanto la detección como la cuantificación de una secuencia específica de una muestra de ácido nucleico diana. El procedimiento puede seguir el principio general de la reacción en cadena de la polimerasa, con la característica

adicional de que el ácido nucleico diana amplificado puede cuantificarse a medida que se va acumulando en la reacción en tiempo real tras cada ciclo de amplificación. Puede haber métodos de cuantificación: (1) el uso de colorantes fluorescentes que se intercalan con ácido nucleico diana bicatenario; y (2) sondas de oligonucleótidos de ADN modificadas que producen fluorescencia cuando se hibridan con un ácido nucleico diana complementario. En el primer método, un colorante de unión a ácido nucleico diana puede unirse a todos los ácidos nucleicos bicatenarios (bc) en la PCR, produciendo la fluorescencia del colorante. Por lo tanto, un aumento en el producto de ácido nucleico durante la PCR puede conducir a un aumento en la intensidad de fluorescencia y puede medirse en cada ciclo, permitiendo así que se cuantifiquen las concentraciones de ácido nucleico. La reacción se puede preparar de manera similar a una reacción de PCR convencional, con la adición de colorante fluorescente de ácido nucleico (bc). La reacción se puede ejecutar en un termociclador, y tras cada ciclo, se pueden medir los niveles de fluorescencia con un detector; el colorante solo puede emitir fluorescencia cuando se una al ácido nucleico (bc) (es decir, el producto de PCR). Con referencia a una dilución patrón, se puede determinar la concentración de ácido nucleico (bc) en la PCR. Los valores obtenidos no pueden tener unidades absolutas asociadas. La comparación de una muestra de ADN/ARN medida con una dilución patrón puede dar una fracción o proporción de la muestra con respecto al patrón, lo que permite comparaciones relativas entre diferentes tejidos o condiciones experimentales. Para garantizar la precisión de la cuantificación, se puede normalizar la expresión de un gen diana con respecto a un gen expresado de forma estable. Esto puede permitir la corrección de posibles diferencias en la cantidad o calidad de ácido nucleico entre muestras. El segundo método puede usar una sonda basada en ARN o ADN específica de la secuencia para cuantificar solo el ácido nucleico que contiene la secuencia de la sonda; por tanto, el uso de la sonda indicadora puede aumentar la especificidad y puede permitir la cuantificación incluso en presencia de alguna amplificación de ácido nucleico no específica. Esto puede permitir la multiplexación (es decir, analizar varios genes en la misma reacción usando sondas específicas con marcadores de diferentes colores), siempre que todos los genes se amplifiquen con una eficiencia similar. Este método puede llevarse a cabo con una sonda basada en ácido nucleico con un indicador fluorescente (por ejemplo, 6-carboxifluoresceína) en un extremo y un inhibidor (por ejemplo, 6-carboxitetrametilrodamina) de la fluorescencia en el extremo opuesto de la sonda. La proximidad del indicador al desactivador puede evitar la detección de su fluorescencia. La rotura de la sonda por la actividad exonucleasa de 5’ a 3' de una polimerasa (por ejemplo, la polimerasa Taq) puede romper la proximidad entre el indicador el desactivador y, por lo tanto, puede permitir que no se apague la emisión de fluorescencia, que se puede detectar. Un aumento en el producto diana de la sonda indicadora en cada ciclo de PCR puede producir un aumento proporcional de la fluorescencia debido a la rotura de la sonda y a la liberación del indicador.

La reacción se puede preparar de manera similar a una reacción de PCR convencional, y se puede añadir la sonda indicadora. A medida que comienza la reacción, durante la etapa de hibridación de la PCR, tanto la sonda como los cebadores pueden hibridarse al ácido nucleico diana. La polimerización de una nueva cadena de ADN se puede iniciar a partir de los cebadores, y una vez que la polimerasa alcanza a la sonda, su exonucleasa 5'-3' puede degradar la sonda, separando físicamente el indicador fluorescente del desactivador, lo que produce un aumento de la fluorescencia. La fluorescencia se puede detectar y medir en un termociclador de PCR en tiempo real, y el aumento geométrico de la fluorescencia puede corresponder al aumento exponencial del producto que se usa para determinar el ciclo umbral en cada reacción. Las concentraciones relativas de ADN presentes durante la fase exponencial de la reacción se pueden determinar rep`representando la fluorescencia frente al número de ciclos en una escala logarítmica (por lo que una cantidad que aumenta exponencialmente puede dar una línea recta). Se puede determinar un umbral para la detección de la fluorescencia por encima del fondo. El ciclo en el que la fluorescencia de una muestra atraviesa el umbral se puede denominar umbral de ciclo, Ct. Dado que la cantidad de ADN puede duplicarse en cada ciclo durante la fase exponencial, se pueden calcular cantidades relativas de ADN (por ejemplo, una muestra con un Ct de 3 ciclos antes que otro tiene 23 = 8 veces más plantilla). Las cantidades de ácido nucleico (por ejemplo, ARN o ADN) se pueden determinar comparando los resultados con una curva patrón producida por una PCR en tiempo real de diluciones en serie (por ejemplo, sin diluir, 1:4, 1:16, 1:64) de un cantidad conocida de ácido nucleico. La reacción de QPCR puede implicar un enfoque de fluoróforo doble que aprovecha la transferencia de energía de resonancia de fluorescencia (FRET), (por ejemplo, sondas de hibridación LIGHTCYCLER, en donde dos sondas oligonucleotídicas pueden hibridarse con el amplicón). Los oligonucleótidos pueden diseñarse para hibridarse en una orientación de cabeza a cola con los fluoróforos separados a una distancia que sea compatible con la transferencia eficaz de energía. Otros ejemplos de oligonucleótidos marcados que están estructurados para emitir una señal cuando se unen a un ácido nucleico o que se incorporan a un producto de extensión incluyen: sondas SCORPIONS, cebadores Sunrise (o AMPLIFLOUR) y cebadores LUX y sondas MOLECULAR BEACONS. La reacción de QPCR puede usar la metodología fluorescente Taqman y un instrumento capaz de medir la fluorescencia en tiempo real (por ejemplo, ABI Prism 7700 Sequence Detector). La reacción de Taqman puede usar una sonda de hibridación marcada con dos colorantes fluorescentes diferentes. Un colorante puede ser un colorante indicador (6-carboxifluoresceína), el otro puede ser un colorante desactivador (6-carboxi-tetrametilrodamina). Cuando la sonda está intacta, puede producirse una transferencia de energía fluorescente, y la emisión fluorescente del colorante indicador puede ser absorbida por el colorante desactivador. Durante la fase de extensión del ciclo de PCR, la sonda de hibridación fluorescente se puede escindir mediante actividad nucleolítica 5'-3’ de la ADN polimerasa. En la escisión de la sonda, la emisión de colorante indicador ya no puede transferirse eficazmente al colorante desactivador, lo que produce un aumento de los espectros de emisión fluorescente del colorante indicador. Se puede usar cualquier método de cuantificación de ácido nucleico, incluyendo los métodos en tiempo real o los métodos de detección de un solo punto para cuantificar la cantidad de ácido nucleico en la muestra. La detección se puede realizar con varias metodologías diferentes (por ejemplo, tinción, hibridación con una sonda marcada; incorporación de cebadores biotinilados seguidos de detección de conjugado de

avidina-enzima; incorporación de trifosfatos desoxinucleotídicos marcados con 32P, tales como dCTP o dATP, en el segmento amplificado. La cuantificación puede o no incluir un etapa de amplificación. La cuantificación no puede ser experimental.

Micromatriz

Los métodos de detección de variantes de secuencia pueden comprender secuenciar y/o detectar las variantes usando una micromatriz. Las micromatrices pueden usarse para determinar el nivel de expresión de una pluralidad de genes en una muestra de ácido nucleico. Las micromatrices se pueden usar para determinar la identidad de secuencia de una pluralidad de secuencias de una muestra de ácido nucleico.

Una micromatriz puede comprender un sustrato. Los sustratos pueden incluir, pero sin limitación, vidrio y vidrio modificado o funcionalizado, plásticos (incluidos acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™ y similares), polisacáridos, nailon y nitrocelulosa, resinas, sílice o materiales a base de sílice, incluidos el silicio y el silicio modificado, carbono, metales, vidrios inorgánicos y plásticos.

Las micromatrices pueden comprender una pluralidad de sondas polinucleotídicas. Una micromatriz puede comprender aproximadamente 1, 10, 100, 1000, 5000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 110000, 120000 o más sondas.

Las sondas pueden tener al menos 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 nucleótidos o más de longitud.

En algunos aspectos, las sondas pueden comprender información de secuencia para un conjunto específico de genes y/o especies. Una sonda puede ser complementaria a una secuencia de ácido nucleico que codifica una proteína hospedadora. Una sonda puede ser complementaria a una secuencia de ácido nucleico no codificante. Una sonda puede ser complementaria a una secuencia de ADN. Una sonda puede ser complementaria a una secuencia de ARN.

Las sondas se pueden inmovilizar en una micromatriz. La inmovilización de polinucleótidos en un sustrato sólido se puede realizar mediante la síntesis directa (por ejemplo, síntesis fotolitográfica) de polinucleótidos en un sustrato sólido o por inmovilización (manchado) de polinucleótidos sintetizados previamente en regiones predeterminadas de un sustrato sólido. Los polinucleótidos se pueden inmovilizar en un sustrato de micromatriz activando una superficie de un sustrato sólido con un grupo funcional nucleófilo (por ejemplo, un grupo amino), acoplando biomoléculas (por ejemplo, polinucleótidos) activadas con un buen grupo saliente al sustrato sólido activado superficialmente, y eliminando los reactivos no reaccionados. Las sondas se pueden inmovilizar en una perla conjugada además a través de una unión covalente o iónica a un soporte sólido. Las sondas se pueden inmovilizar sobre un sustrato usando una película específica que tenga una baja conductividad y una temperatura de fusión baja, es decir, una película de oro. Una radiación electromagnética aplicada puede fundirse y puede ablacionar la película en el sitio de impacto. La película puede estar en contacto con una dispersión coloidal y, al fundirse, puede generar un flujo de convección en el sitio de reacción, lo que conduce a la adhesión de una partícula insoluble en la dispersión a un sitio fundido específicamente.

Una micromatriz puede analizar una muestra de ácido nucleico que comprenda ácidos nucleicos de identidad desconocida (por ejemplo, muestra de prueba) comparando la muestra de ácido nucleico de identidad desconocida con una muestra de referencia. Se puede preparar una muestra de ácido nucleico a partir de ADN (por ejemplo, ADN aislado, ADN genómico, ADN extracromasómico). Se puede preparar una muestra de ácido nucleico a partir de ARN. El ARN se puede transcribir inversamente en ADN con un cebador específico del gen o un cebador universal. El ADN transcrito inversamente (por ejemplo, ADNc) puede tratarse con Rnasa o una base (por ejemplo, NaOH) para hidrolizar el ARN. El ADNc puede marcarse con un colorante (por ejemplo, Cy3, Cy5) con química de *N*-hidroxisuccinimida o químicas de marcaje similares. Los colorantes fluorescentes adecuados pueden incluir varios colorantes comerciales y derivados de colorantes tales como los que se denominan Alexa, Fluoresceína, Rodamina, FAM, TAMRA, Joe, ROX, Texas Red, BODIPY, FITC, Oregon Green, Lisamina y otros. La muestra de referencia se puede marcar con un colorante diferente al de la muestra de prueba.

La muestra de prueba y la muestra de referencia se pueden aplicar a una micromatriz para poner en contacto múltiples puntos simultáneamente. La muestra de prueba y la muestra de referencia se pueden aplicar a la micromatriz en condiciones de hibridación que pueden permitir que los ácidos nucleicos de la muestra de ácido nucleico se unan a una sonda complementaria o en la micromatriz. Se pueden realizar varias etapas de reacción con las moléculas unidas en la micromatriz, incluyendo la exposición de moléculas reactivas unidas a las etapas de lavado. Se puede controlar el progreso o el resultado de la reacción en cada punto (por ejemplo, sonda) en la micromatriz para caracterizar la muestra de ácido nucleico inmovilizada en el chip. En general, el análisis de micromatrices puede requerir un período de incubación que puede variar de minutos a horas. La duración del período de incubación puede depender del ensayo, y puede determinarse por varios factores, tales como el tipo de reactivo, el grado de mezcla, el volumen de la muestra, el número de copias diana y la densidad de la matriz. Durante el período de incubación, los ácidos nucleicos de la muestra de ácido nucleico pueden estar en contacto íntimo con las sondas de micromatrices.

La detección se puede realizar usando un instrumento de barrindo confocal con excitación láser y detección de tubo fotomultiplicador, tal como el ScanArray 3000 proporcionado por GSI Lumonics (Bellerica, MA). Para implementar el método, se pueden usar sistemas de detección fluorescente confocales y no confocales tales como los proporcionados por Axon Instruments (Foster City, CA), Genetic MicroSystems (Santa Clara, CA), Molecular Dynamics (Sunnyvale, CA) y Virtek (Woburn, MA). Los sistemas de detección alternativos pueden incluir sistemas de barrido que usan láseres en estado gaseoso, diodo y sólido, así como aquellos que usan otros varios tipos de fuentes de iluminación, tales como bombillas de xenón y halógenas. Además de los tubos fotomultiplicadores, los detectores pueden incluir cámaras que usan dispositivos de acoplamiento de carga (CCD) y chips complementarios de silicio de óxido de metal (CMOS).

Se puede comparar la proporción de las intensidades de los dos colorantes de la muestra de prueba y la muestra de referencia para cada sonda. La intensidad de la señal detectada procedente de un punto de micromatriz dado puede ser directamente proporcional al grado de hibridación de un ácido nucleico de la muestra a la sonda en un punto dado (por ejemplo, un punto comprende una sonda). El análisis de las intensidades de fluorescencia de las micromatrices hibridadas puede incluir la segmentación de los puntos, la determinación del fondo (y la posible sustracción), la eliminación de puntos malos, seguido de un método de normalización para corregir cualquier ruido restante. Las técnicas de normalización pueden incluir la normalización global en todos los puntos o un subconjunto de los puntos, tal como los genes constitutivos, el desplazamiento prelogarítmico para obtener mejores coincidencias de referencia, o en el caso de dos (o más) hibridaciones de canales que encuentran el mejor ajuste que ayuda a dar una gráfica de M frente a A centrada en torno a M = 0 y/o que ayuda a dar una gráfica del log (Rojo) frente al log (Verde) centrada en torno A LA diagonal con la menor propagación. La gráfica de M frente a A también se puede denominar gráfica de R frente a I, en la que R es una proporción, tal como $R = \log_2(\text{Rojo/Verde})$ e I es una intensidad, tal como I = log VRcd* Verde. La escala, el desplazamiento, los mejores ajustes a través de diagramas de dispersión, etc. pueden ser técnicas usadas para normalizar conjuntos de datos de micromatrices y proporcionar una mejor base para el análisis posterior. La mayoría de estos métodos de normalización pueden tener algunas hipótesis subyacentes (tales como "la mayoría de los genes dentro del estudio no varían mucho").

Ácidos nucleicos marcados y métodos de uso

Visión general

La divulgación proporciona kits, métodos y composiciones para ácidos nucleicos dirigidos a ácido nucleico marcados, como se describe en el presente documento. La **Figura 10** representa un ejemplo de ácido nucleico dirigido a ácido nucleico **1005** de la divulgación. Un ácido nucleico dirigido a ácido nucleico puede comprender una o más secuencias no nativas (por ejemplo, marcadores) **1010**/**1015**. Un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia no nativa **1010**/**1015** en el extremo 3', el extremo 5' o ambos extremos 3' y 5' del ácido nucleico dirigido a ácido nucleico.

En algunos casos, un ácido nucleico dirigido a ácido nucleico puede ser ácido nucleico dirigido a ácido nucleico tal como se describe en el presente documento, y comprender una o más secuencias no nativas, tales como en el extremo 3', el extremo 5' o ambos extremos 3' y 5' del ácido nucleico dirigido a ácido nucleico.

En algunos casos, el ácido nucleico dirigido a ácido nucleico es un ácido nucleico dirigido a ácido nucleico guía doble tal como se describe en el presente documento, y tal como se representa en la **Figura 11**. Un ácido nucleico dirigido a ácido nucleico guía doble puede comprender dos moléculas de ácido nucleico **1105**/**1110**. Un ácido nucleico dirigido a ácido nucleico guía doble puede comprender una pluralidad de secuencias no nativas **1115**/**1120**/**1125**/**1130**. La secuencia no nativa se puede ubicar en el extremo 3', el extremo 5' o ambos extremos 3' y 5' de cada molécula del ácido nucleico dirigido a ácido nucleico. Por ejemplo, la secuencia no nativa puede estar situada en el extremo 3', extremo 5' o en ambos extremos 3' y 5' de la primera molécula de ácido nucleico **1105**. La secuencia no nativa se puede ubicar en el extremo 3', extremo 5' o ambos extremos 3' y 5' de la segunda molécula **1110**. Un ácido nucleico dirigido a ácido nucleico puede comprender una o más secuencias no nativas en cualquiera de las configuraciones enumeradas en la **Figura 11**.

La divulgación proporciona métodos de uso de ácidos nucleicos dirigidos a ácido nucleico marcados. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. En algunos casos, una pluralidad de ácidos nucleicos dirigidos a ácido nucleico marcados pueden ponerse en contacto con una pluralidad de ácidos nucleicos diana. La **Figura 12** representa un método de uso ilustrativo para los ácidos nucleicos dirigidos a ácido nucleico marcados. Un ácido nucleico dirigido a ácido nucleico marcado puede comprender un espaciador **1210** que puede hibridarse con un ácido nucleico diana **1205**. El ácido nucleico dirigido a ácido nucleico puede comprender una secuencia no nativa (por ejemplo, marcador) **1220**. La secuencia no nativa **1220** puede ser una secuencia de unión de proteínas de unión a ARN. En algunos casos, la secuencia no nativa **1220** puede ser una secuencia de unión de proteína de unión a RNA CRISPR. La secuencia no nativa **1220** puede estar unida por una proteína de unión a ARN **1215**. La proteína de unión a ARN **1215** puede comprender una secuencia no nativa **1225** (por ejemplo, una fusión, es decir, la proteína de unión a ARN **1215** puede ser un polipéptido de fusión). La secuencia no nativa (por ejemplo, fusión) **1225** puede alterar la transcripción del ácido

nucleico diana y/o un ácido nucleico exógeno. La secuencia no nativa (por ejemplo, fusión) **1225** puede comprender una primera parte de un sistema dividido.

En algunos aspectos, un segundo ácido nucleico dirigido a ácido nucleico, que comprende un segundo espaciador **1240**, que puede hibridarse con un segundo ácido nucleico diana **1245**, puede comprender una segunda secuencia no nativa (por ejemplo, marcador) **1250**. La segunda secuencia no nativa (por ejemplo, marcador) **1250** puede ser una secuencia de unión de proteínas de unión a ARN. La segunda secuencia no nativa (por ejemplo, marcador) **1250** puede ser una secuencia de unión de proteínas de unión a ARN CRISPR. La segunda secuencia no nativa **1250** puede estar unida por una proteína de unión a ARN **1235**. La proteína de unión a ARN puede comprender una secuencia no nativa **1230** (por ejemplo, fusión, es decir, la proteína de unión a ARN **1235** puede ser una fusión). La secuencia no nativa **1230** (por ejemplo, fusión) puede ser una segunda parte de un sistema dividido.

En algunos casos, la primera parte del sistema dividido **1225** y la segunda parte del sistema dividido **1230** pueden estar cercanas entre sí en el espacio, de manera que la primera parte del sistema dividido **1225** y la segunda parte del sistema dividido **1230** interactúan **1255** para formar un sistema dividido activo **1260**. Un sistema dividido activo **1260** puede referirse a un sistema no dividido, en donde la primera parte y la segunda parte forman una pieza completa del sistema dividido. La activación del sistema dividido puede indicar que dos ácidos nucleicos diana **1205**/**1245** están muy juntos entre sí en el espacio.

Métodos

La divulgación proporciona métodos para poner en contacto un ácido nucleico diana con un complejo que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico, e introducir una o más proteínas efectoras, en donde la una o más proteínas efectoras comprenden una secuencia no nativa y pueden unirse al ácido nucleico dirigido a ácido nucleico modificado. Una proteína efectora puede referirse a cualquier proteína con un efecto funcional. Por ejemplo, una proteína efectora puede comprender actividad enzimática, remodelar moléculas biológicas (por ejemplo, chaperonas plegables), ser una proteína de armazón y/o unirse a una molécula pequeña o metabolito. La proteína efectora puede modificar el ácido nucleico diana (por ejemplo, por escisión, modificación enzimática, modificación de la transcripción). Los métodos de la divulgación proporcionan el uso de las composiciones de la divulgación tales como biosensores. Por ejemplo, los complejos (por ejemplo, los que comprenden un ácido nucleico dirigido a ácido nucleico modificado, un polipéptido dirigido a un sitio y/o una proteína efectora) se pueden usar para controlar los eventos de movilidad genética, detectar cuándo las secuencias están muy juntas en un espacio tridimensional, y alterar condicionalmente la transcripción.

Evento de movilidad genética

La divulgación proporciona métodos para determinar la existencia de un evento de movilidad genética. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. Un evento de movilidad genética puede comprender, por ejemplo, una translocación, una recombinación, una integración, una transposición, un evento horizontal de transferencia génica, una transformación, una transducción, una conjugación, un evento de conversión génica, una duplicación, una translocación, una inversión, una eliminación, una sustitución o cualquier combinación de los mismos.

Un evento de movilidad genética puede comprender una recombinación entre genes. La recombinación puede conducir a productos génicos perjudiciales (por ejemplo, la recombinación BCR-ABL que puede contribuir al cáncer de mama). La recombinación puede incluir, por ejemplo, recombinación homóloga, recombinación no homóloga (por ejemplo, unión de extremos no homólogos) y la recombinación V(D)J. La recombinación puede referirse al cruce cromosómico. La recombinación puede ocurrir durante la profase I de la meiosis (por ejemplo, sinapsis). La recombinación puede comprender la rotura bicatenaria de las cadenas de ADN de ácido nucleico, seguida de la formación de una unión de holliday por recombinasas que pueden catalizar el intercambio de las cadenas de ADN.

Los eventos de movilidad genética pueden causar enfermedades. Por ejemplo, la leucemia mielógena crónica puede ser el resultado de un evento de movilidad genética. La translocación entre los cromosomas 9 y 22 puede producir un gen de fusión BCR-Abl1, que puede provocar el alargamiento de un cromosoma (por ejemplo, 9) y el acortamiento de otro cromosoma (por ejemplo, 22, es decir, el cromosoma Filadelfia). La translocación de BCR-Abl1 puede conducir a la producción de una proteína de fusión BCR-Abl que puede interactuar con los receptores (por ejemplo, el receptor de interleucina-3) para potenciar la división celular, lo que conduce a la leucemia mielógena crónica (LMC). Otros eventos de movilidad genética ilustrativos no limitantes incluyen BRD3-NUT, BRD4-NUT, KIAA1549-BRAF, FIG/GOPC-ROS1, ETV6-NTRK3, BCAS4-BCAS3, TBL1XR1-RGS17, ODZ4-NRG1, MALAT1-TFEB, APSCR1-TFE3, PRCC-TFE3, CLTC-TFE3, NONO-TFE3, SFPQ-TFE3, ETV6-NRTK3, EML4-ALK, EWSR1-ATF1, MN1-ETV6, CTNNB1-PLAG1, LIFR-PLAG1, TCEA1-PLAG1, FGFr1-PLAG1, CHCHD7-PLAG1, HMGA2-FHIT, HMGA-NFIB, CRTC1-MAMl2, CRCT3-MAML2, EWSR1-POUF5F1, TMPRSS1-ERG, TMPRSS2-ETV4, TMPRSS2-ETV5, HNRNPA2B1 -ETV1, HERV-K-ETV1, C150RF21-ETV1, SLC45A3-ETV1, SLC45A3-ETV5, SLC45A3-ELK4, KLK2-ETV4, CANT1-ETV4, RET-PTC1/CCDC6, RET-PTC2/PRKAR1A, RET-PTC3,4/NCOA4, RET-PTC5/GOLGA5, RET-PTC6/TRIM24, RET-PTC7/TRIM33, RET-PTC8/KTN1, RET-PTC9/RFG9, RET-PTCM1, TFG-NTRK1, TPM3-NRTK1,

TPR-NRTK1, RET-D10S170, ELKS-RET, HOOKS3-RET, RFP-RET, AKAP9-BRAF y PAX8-PPARG.

Las enfermedades que pueden ser causadas por eventos de movilidad genética pueden incluir la enfermedad de Charcot-Marie-Tooth de tipo 1A (CMT1A), nefronofisis juvenil (NPH), ictiosis ligada al cromosoma X, deficiencia de hormona de crecimiento familiar de tipo 1 A, distrofia muscular fascioescapulohumeral (FSHD), α-talasemia, hemofilia A, síndrome de Hunter (es decir, mucopolisacaridosis II), distrofia muscular de Emery-Dreifuss, Síndrome de hemoglobina Lepore, deficiencia de esteroides 21-hidroxilasa, hiperaldosteronismo supresor de glucocorticoides (GSH), daltonismo (por ejemplo, dicromato visual), atrofia muscular espinal (AME), cáncer, leucemia linfoblástica aguda de linfocitos T (LLA-T), carcinoma agresivo de línea media, Astrocitoma, Carcinoma de mama secretor, Cáncer de mama, Carcinoma de riñón, Nefroma mesoblástico, Adenocarcinoma de pulmón, Melanoma, Meningioma, adenoma pleomórfico, cáncer mucoepidermoide, Carcinoma de próstata, carcinoma de tiroides y leucemia promielocítica aguda.

Los métodos de la divulgación proporcionan la determinación de la existencia de un evento de movilidad genética en donde un ácido nucleico diana puede ponerse en contacto con dos complejos, comprendiendo cada complejo un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico modificado, y se pueden introducir dos o más proteínas efectoras, en donde las dos o más proteínas efectoras pueden unirse a los ácidos nucleicos dirigidos a ácido nucleico modificados, en donde una de las dos o más proteínas efectoras comprende una secuencia no nativa que es una primera pieza de un sistema dividido y una de las dos o más proteínas efectoras comprende una secuencia no nativa que es una segunda pieza del sistema dividido. Un sistema dividido puede referirse a un complejo de proteínas compuesto por dos o más fragmentos de proteínas que individualmente no son fluorescentes, pero que, cuando se forman en un complejo, producen un complejo de proteínas fluorescentes funcional (es decir, que produce fluorescencia). Los fragmentos de proteínas individuales de un sistema dividido (por ejemplo, proteína fluorescente dividida) pueden denominarse "fragmentos de complementariedad" o "fragmentos complementarios". Los fragmentos complementarios que pueden ensamblarse espontáneamente en un complejo proteico fluorescente funcional se conocen como fragmentos complementarios de autocomplementación, autoensamblaje o asociación espontánea. Por ejemplo, un sistema dividido puede comprender GFP. En un sistema dividido de GFP, los fragmentos complementarios se derivan de la estructura tridimensional de GFP, que incluye once cadenas beta externas antiparalelas y una cadena alfa interna. Un primer fragmento puede comprender una de las once cadenas beta de la molécula de GFP (por ejemplo, GFP S11), y un segundo fragmento puede comprender las cadenas restantes (por ejemplo, GFP S1-10).

Antes del evento de movilidad genética, la secuencia de ácido nucleico diana que puede ser dirigida por un complejo puede estar muy separada de la secuencia de ácido nucleico diana que puede ser dirigida por otra secuencia. La distancia entre las dos secuencias de ácido nucleico diana puede comprender al menos aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más Kb. La distancia entre las dos secuencias de ácido nucleico diana puede comprender, como máximo, aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más Kb. Las dos secuencias de ácido nucleico diana pueden ubicarse en diferentes cromosomas. Las dos secuencias de ácido nucleico diana pueden ubicarse en el mismo cromosoma.

Antes del evento de movilidad genética, las proteínas efectoras que comprenden partes del sistema dividido pueden no ser capaces de interactuar entre sí (por ejemplo, el sistema dividido puede estar inactivo). Tras el evento de movilidad genética, la secuencia de ácido nucleico diana que puede ser dirigida por un complejo puede ubicarse muy cerca de la secuencia de ácido nucleico diana que puede ser dirigida por el otro complejo. Tras el evento de movilidad genética, las proteínas efectoras que comprenden partes del sistema dividido pueden interactuar entre sí, activando así el sistema dividido.

El sistema dividido activado puede indicar la existencia del evento de movilidad genética. Por ejemplo, si el sistema dividido activado es un sistema dividido de proteína fluorescente, entonces, antes del evento de movilidad genética, puede no detectarse fluorescencia en la muestra. En algunos casos, los niveles de fluorescencia del sistema dividido inactivo (por ejemplo, niveles de fondo) pueden ser 0,01, 0,05, 0,1, 0,5, 1, 1,5, 2, 2,5, 3, 3,5 o más veces menos fluorescentes en comparación con una muestra de control (por ejemplo, célula) que no comprende el sistema dividido. En algunos casos, los niveles de fluorescencia del sistema dividido inactivo (por ejemplo, nivel de fondo) pueden ser 0,01, 0,05, 0,1, 0,5, 1, 1,5, 2, 2,5, 3, 3,5 o más veces más fluorescentes que una muestra de control (por ejemplo, célula) que no comprende el sistema dividido.

Tras el evento de movilidad genética, las dos piezas divididas pueden unirse para formar una proteína fluorescente activa, y se puede detectar fluorescencia en la muestra. Un sistema dividido activo puede producir un aumento de fluorescencia de al menos aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces. Un sistema dividido activo puede producir un aumento de la fluorescencia de 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces.

La detección de una movilidad genética se puede usar para genotipar un sujeto (por ejemplo, un paciente). Un genotipo puede ser indicativo de una enfermedad. La detección de un evento de movilidad genética se puede usar para diagnosticar un sujeto. La información genética y de diagnóstico obtenida de los métodos descritos en el presente documento se puede comunicar a un sujeto. La información genética y de diagnóstico obtenida a partir de los métodos descritos en el presente documento se puede usar para desarrollar un plan de tratamiento específico del sujeto. Por ejemplo, si los datos obtenidos a partir de los métodos de la divulgación indican que un paciente tiene un genotipo que

le hace resistente a un régimen terapéutico en particular, se puede hacer un nuevo plan de tratamiento para el sujeto.

Alteración de la transcripción

Los métodos de la divulgación pueden proporcionar la alteración de la transcripción de un ácido nucleico. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. El método proporciona el contacto de un ácido nucleico diana con dos complejos, comprendiendo cada complejo un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico modificado,y la introducción de dos o más proteínas efectoras, en donde las dos o más proteínas efectoras pueden unirse a los ácidos nucleicos dirigidos a ácido nucleico modificados, en donde una de las dos o más proteínas efectoras comprende una secuencia no nativa que es una primera pieza de un sistema de factor de transcripción dividido y una de las dos o más proteínas efectoras comprende una secuencia no nativa que es una segunda pieza del sistema de factor de transcripción dividido, y en donde una interacción entre la primera pieza y la segunda pieza del sistema dividido forma un factor de transcripción que altera la transcripción del ácido nucleico.

El factor de transcripción puede alterar los niveles de transcripción de un ácido nucleico y/o un ácido nucleico diana. La transcripción alterada puede incluir niveles de transcripción aumentados y/o niveles de transcripción disminuidos. Un factor de transcripción puede alterar los niveles de transcripción más de 2 veces, 3 veces, 5 veces, 10 veces, 50 veces, 100 veces, 1000 veces o más superior o inferior que los niveles de transcripción no alterados. Un factor de transcripción puede alterar los niveles de transcripción menos de 2 veces, 3 veces, 5 veces, 10 veces, 50 veces, 100 veces, 1000 veces o más superior o inferior que los niveles de transcripción no alterados.

El factor de transcripción puede alterar la transcripción de un ácido nucleico diana y/o un ácido nucleico exógeno. Un ácido nucleico diana puede ser el ácido nucleico que está en contacto con el complejo que comprende el polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico. Un ácido nucleico exógeno puede comprender un polinucleótido donante, un plásmido y/o un ácido nucleico diana.

Un ácido nucleico exógeno puede comprender un polinucleótido que codifica genes implicados en la apóptosis. Los genes adecuados implicados en la apóptosis pueden incluir factor de necrosis tumoral (TNF), TNF-R1, TNF-R2, dominio de muerte asociado al receptor de TNF (TRADD), receptor Fas y ligando Fas, caspasas (por ejemplo, caspasa-3, caspasa-8, caspasa-10), APAF-1, FADD y factor inductor de apóptosis (AIF). Un ácido nucleico exógeno puede comprender un polinucleótido que codifica genes que producen la lisis celular. Los genes adecuados pueden incluir la proteína de muerte de Adenovirus (ADP), defensinas, péptidos líticos permeabilizantes de membrana derivados de c-FLIP, procaspasas, péptidos penetrantes de células, por ejemplo, VIH TAT. Un ácido nucleico exógeno puede comprender un polinucleótido que codifica un antígeno que puede producir el reclutamiento de células inmunitarias en la ubicación celular (por ejemplo, péptidos de clase MHC). Un ácido nucleico exógeno puede comprender un polinucleótido que codifica un ácido nucleico dirigido a ácido nucleico que se dirige a secuencias que ocurren muchas veces dentro del genoma (por ejemplo, microsatélites, repeticiones en tándem), produciendo una fragmentación del genoma a gran escala y la muerte celular.

Modificación del ácido nucleico diana

La divulgación proporciona métodos para modificar un ácido nucleico diana usando el ácido nucleico dirigido a ácido nucleico de la divulgación. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. Por ejemplo, un ácido nucleico diana puede ponerse en contacto con un complejo que comprenda un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y una o más proteínas efectoras, en donde la una o más proteínas efectoras comprenden una secuencia no nativa y pueden unirse al ácido nucleico dirigido a ácido nucleico modificado. La secuencia no nativa puede conferir una actividad enzimática y/o la actividad de transcripción de la proteína efectora puede modificar el ácido nucleico diana. Por ejemplo, si la proteína efectora comprende una secuencia no nativa correspondiente a una metiltransferasa, entonces la metiltransferasa puede ser capaz de metilar el ácido nucleico diana. La modificación del ácido nucleico diana puede ocurrir al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos lejos del extremo 5' o 3 'del ácido nucleico diana. La modificación del ácido nucleico diana puede ocurrir, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos lejos del extremo 5' o 3' del ácido nucleico diana. La modificación puede ocurrir en un ácido nucleico separado que no comprenda el ácido nucleico diana (por ejemplo, otro cromosoma).

Las modificaciones ilustrativas pueden comprender metilación, desmetilación, acetilación, desacetilación, ubiquitinación, desubiquitinación, desaminación, alquilación, despurinación, oxidación, formación de dímeros de pirimidina, transposición, recombinación, alargamiento de cadena, ligamiento, glicosilación, fosforilación, desfosforilación, adenilación, desadenilación, SUMOilación, deSUMOilación, ribosilación, desribosilación, miristoilación, remodelación, escisión, oxidorreducción, hidrolación e isomerización.

Determinación de un genotipo y tratamiento

La divulgación proporciona métodos para tratar una enfermedad usando el ácido nucleico dirigido a ácido nucleico de la divulgación. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. Por ejemplo, usando el sistema dividido descrito en el presente documento, la presencia de dos o más ácidos nucleicos diana juntos en el espacio (por ejemplo, en un evento de movilidad genética, en la estructura de la cromatina o en un ácido nucleico lineal) puede ser indicativo de un genotipo (por ejemplo, de un sujeto). Un genotipo puede referirse a la presencia o ausencia de una determinada secuencia de ácido nucleico, un polimorfismo de nucleótidos (es decir, un polimorfismo de un solo nucleótido o un polimorfismo de múltiples nucleótidos), una variante alélica o cualquier otra indicación de la secuencia de un ácido nucleico. El genotipo puede indicar si un paciente padece una enfermedad y/o está predispuesto a contraer una enfermedad.

La determinación de un genotipo puede incluir, por ejemplo, determinar si un sujeto comprende una secuencia mutante (por ejemplo, una secuencia de ácido nucleico que comprende una mutación). En algunos casos, un primer ácido nucleico dirigido a ácido nucleico que comprende los componentes apropiados tal como se describe en el presente documento para comprender una primera parte de un sistema dividido puede diseñarse para dirigirse a una región cerca de una secuencia mutante predicha. En algunos casos, un segundo ácido nucleico dirigido a ácido nucleico que comprende los componentes apropiados tal como se describe en el presente documento para comprender una segunda parte del sistema dividido puede diseñarse para dirigirse a una región que comprenda la secuencia mutante predicha. Si la secuencia mutante existe, el segundo ácido nucleico dirigido a ácido nucleico puede unirse a ella, y las dos partes del sistema dividido pueden interactuar. La interacción puede generar una señal que puede ser indicativa de la presencia de una secuencia mutante.

Un genotipo puede ser identificado por un biomarcador. A biomarcador puede ser indicativo de cualquier proceso fisiológico. Un biomarcador puede servir como indicador de la eficacia de un tratamiento (por ejemplo, tratamiento con fármacos). Un biomarcador puede ser un ácido nucleico, un polipéptido, un analito, un soluto, una molécula pequeña, un ion, un átomo, una modificación de un ácido nucleico y/o polipéptido y/o un producto de degradación. Un biomarcador puede referirse a niveles de expresión relativos de un ácido nucleico y/o un polipéptido.

Se puede identificar un plan de tratamiento específico del sujeto a partir de la determinación del genotipo del sujeto usando los métodos de la divulgación. Por ejemplo, si un sujeto comprende un cierto genotipo que se sabe que no responde a una terapia en particular, entonces el sujeto puede ser tratado con una terapia diferente. La determinación del genotipo puede permitir que un sujeto sea seleccionado o excluido para un ensayo clínico.

La determinación del genotipo puede ser comunicada por un cuidador a un sujeto (por ejemplo, por un médico a un paciente o por una persona que realiza el análisis del genotipo a un cliente). La comunicación puede ocurrir en persona (por ejemplo, en un consultorio médico), por teléfono, por escrito o electrónicamente. La comunicación puede informar además al sujeto de un régimen de tratamiento específico del sujeto, determinado a partir del genotipo del sujeto.

El método se puede realizar más de una vez (por ejemplo, iterativamente) en un sujeto. Por ejemplo, se puede determinar el genotipo de un sujeto, se puede prescribir un curso de tratamiento para el sujeto, se puede determinar nuevamente el genotipo del sujeto. Los dos genotipos se pueden comparar para determinar la efectividad del curso del tratamiento. El plan de tratamiento puede modificarse en función de la comparación de los genotipos.

Localización de secuencias en el espacio tridimensional

En algunos casos, la divulgación proporciona un método para determinar la ubicación de secuencias en un espacio tridimensional en una célula. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento. La determinación de la organización tridimensional de la cromatina y el ácido nucleico puede ser importante para comprender la regulación génica, tal como la activación de la transcripción y/o la represión de genes. En algunos casos, el método comprende poner en contacto un ácido nucleico diana con dos complejos, en donde cada complejo se une a su ácido nucleico diana afín. Los complejos pueden comprender un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico de la divulgación. Se pueden introducir dos o más proteínas efectoras, en donde cada una de las dos o más proteínas efectoras se une a un complejo. Las proteínas efectoras pueden ser similares al sistema dividido descrito anteriormente, en donde cada proteína efectora puede comprender un fragmento inactivo de un polipéptido completo. Cuando las proteínas efectoras están muy separadas en el espacio, las proteínas efectoras están inactivas (por ejemplo, no se detecta señal). Cuando las proteínas efectoras están lo suficientemente cerca en el espacio para interactuar, pueden formar un polipéptido activo detectable.

Las proteínas efectoras pueden ser parte de un sistema de marcador de afinidad dividido. En un sistema de marcador de afinidad dividido, los dos fragmentos polipeptídicos inactivos del sistema pueden corresponder a dos fragmentos inactivos de un marcador de afinidad. Cuando los dos fragmentos se unen, se restablece el marcador de afinidad completo, de modo que el marcador de afinidad puede ser detectado por un agente de unión. Un agente de unión puede referirse a una molécula que puede unirse y purificar el marcador de afinidad. Los ejemplos de agentes de unión

pueden incluir anticuerpos, perlas conjugadas con anticuerpos y perlas conjugadas de molécula pequeña.

La introducción de los complejos y polipéptidos de la divulgación puede ocurrir por infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares.

Las células pueden cultivarse con los complejos durante al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más días. Las células pueden cultivarse con los complejos, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más días. Tras un período de tiempo apropiado (por ejemplo, un período de tiempo para permitir que los complejos se unan a su ácido nucleico diana), las células se pueden lisar.

Las células se pueden entrecruzar antes del lisado. Las células fijas o entrecruzadas pueden estabilizar los complejos de proteína-ADN en la célula. Los fijadores y agentes de entrecruzamiento adecuados pueden incluir, formaldehído, glutaraldehído, fijadores a base de etanol, fijadores a base de metanol, acetona, ácido acético, tetraóxido de osmio, dicromato de potasio, ácido crómico, permanganato de potasio, mercuriales, picratos, formol, paraformaldehído, agentes de entrecruzamiento de NHS-éster reactivos con amina tales como bis[sulfosuccinimidil]suberato (BS3), 3,3'-ditiobis[sulfosuccinimidilpropionato] (DTSSP), bis[sulfosuccinimidilsuccinato de etilenglicol (sulfo-EGS), glutarato de disuccinimidilo (DSG), propionato de ditiobis[succinimidilo] (DSP), suberato de disuccinimidilo (DSS), bis[succinimidilsuccinato] de etilenglicol (EGS), agentes de entrecruzamiento de NHS-éster/diazirina tales como NHS-diazirina, NHS-LC-diazirina, NHS-SS-diazirina, sulfo-NHS-diazirina, sulfo-NHS-LC-diazirina y sulfo-NHS-SS-diazirina.

Las células lisadas pueden ponerse en contacto con un agente de unión (por ejemplo, un anticuerpo) que se dirija a unirse al marcador de afinidad. El contacto puede ocurrir en un tubo de ensayo. El contacto puede ocurrir en un entorno cromatográfico (por ejemplo, una columna de cromatografía de afinidad). El contacto con el agente aglutinante puede ocurrir durante al menos 1 minuto, 5 minutos, 10 minutos, 15 minutos, 20 minutos, 25 minutos, 30 minutos, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 15 horas, 20 horas, 25 horas, 30 horas, 35 horas, 40 horas, 45 o más horas. El contacto con el agente aglutinante puede ocurrir, como máximo, durante 1 minuto, 5 minutos, 10 minutos, 15 minutos, 20 minutos, 25 minutos, 30 minutos, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 15 horas, 20 horas, 25 horas, 30 horas, 35 horas, 40 horas, 45 o más horas. En algunos casos, el contacto con un agente de unión ocurre antes de la lisis celular.

Los complejos se pueden purificar con el agente aglutinante. Los complejos purificados pueden someterse a técnicas de purificación de ácido nucleico para separar el ácido nucleico diana de los complejos. Las técnicas de purificación de ácido nucleico pueden incluir separación en columna de centrifugación, precipitación y electroforesis.

El ácido nucleico (por ejemplo, el ácido nucleico que compre el ácido nucleico diana) puede someterse a metodologías de secuenciación. El ácido nucleico puede prepararse para el análisis de secuenciación mediante ligadura de uno o más adaptadores. Los ácidos nucleicos secuenciados se pueden analizar para identificar polimorfismos, diagnosticar una enfermedad, determinar un curso de tratamiento para una enfermedad y/o determinar la estructura tridimensional del genoma.

Ácidos nucleicos dirigidos a ácido nucleico marcados con enlazadores

La divulgación proporciona composiciones y métodos para generar y usar ácidos nucleicos dirigidos a ácido nucleico marcados. La **Figura 13A** representa un ejemplo de ácido nucleico dirigido a ácido nucleico sin marcar. Un ácido nucleico dirigido a ácido nucleico sin marcar puede comprender un protoespaciador (PS), una repetición CRISPR mínima (MCR), un conector guía sencillo (SGC), una secuencia mínima de ARNcrtra (MtS), una secuencia de ARNcrtra de 3' (3TS) y extensión de ARNcrtra (TE). Un ácido nucleico dirigido a ácido nucleico marcado que comprende un enlazador puede referirse a cualquiera de los ácidos nucleicos dirigidos a ácido nucleico descritos en el presente documento y que comprende un enlazador en el extremo 5', el extremo 3' o tanto el extremo 5' como 3' del ácido nucleico dirigido a ácido nucleico.

Un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia no nativa como se representa en la **Figura 13B**. La secuencia no nativa se puede denominar marcador. El marcador puede fusionarse con el extremo 5', el extremo 3' o tanto el extremo 5' como 3' del ácido nucleico dirigido a ácido nucleico. La secuencia no nativa puede comprender una secuencia de unión para una proteína de unión a ARN. La proteína de unión a ARN puede ser Csy4. La secuencia no nativa se puede fusionar con la secuencia del protoespaciador del ácido nucleico dirigido a ácido nucleico.

La fusión no nativa se puede separar del ácido nucleico dirigido a ácido nucleico mediante un enlazador. La **Figura 14** representa un enlazador ilustrativo (por ejemplo, enlazador Tag), que separa la secuencia no nativa (por ejemplo, horquilla Csy4) del protoespaciador del ácido nucleico dirigido a ácido nucleico. La secuencia enlazadora puede ser complementaria al ácido nucleico diana. La secuencia enlazadora puede comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más emparejamientos erróneos con el ácido nucleico diana. La secuencia enlazadora puede comprender, como

máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más emparejamientos erróneos con el ácido nucleico diana. En algunos casos, cuanto menos emparejamientos erróneos haya entre el enlazador y el ácido nucleico diana, mejor será la eficacia de escisión del complejo de Cas9:ácido nucleico dirigido a ácido nucleico.

La secuencia enlazadora puede ser de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60 o más nucleótidos de longitud. La secuencia enlazadora puede ser, como máximo, de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60 o más nucleótidos de longitud.

Agentes de dirección genéticos multiplexados

Visión general

La presente divulgación describe métodos, composiciones, sistemas y/o kits para la ingeniería genómica multiplexada. En algunos aspectos de la divulgación, un polipéptido dirigido a un sitio puede comprender un ácido nucleico dirigido a ácido nucleico, formando así un complejo. El complejo puede ponerse en contacto con un ácido nucleico diana. El ácido nucleico diana puede ser escindido y/o modificado por el complejo. Los métodos, las composiciones, los sistemas y/o los kits de la divulgación pueden ser útiles para modificar múltiples ácidos nucleicos diana de forma rápida, eficaz y/o simultánea. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

La **FIGURA 15** representa un método ilustrativo de la divulgación. Un ácido nucleico (por ejemplo, un ácido nucleico dirigido a ácido nucleico) **1505** puede fusionarse con una secuencia no nativa (por ejemplo, una fracción, una secuencia de unión a endorribonucleasa, ribozima) **1510**, formando así un módulo **1512** de ácido nucleico. El módulo **1512** de ácido nucleico (por ejemplo, que comprende el ácido nucleico fusionado a una secuencia no nativa) puede conjugarse en tándem, formando así un agente de dirección genético multiplexado (por ejemplo, polimódulo, por ejemplo, matriz) **1511**. El agente de dirección genético multiplexado **1511** puede comprender ARN. El agente de dirección genético multiplexado puede ponerse en contacto **1515** con una o más endorribonucleasas **1520**. Las endorribonucleasas pueden unirse a la secuencia no nativa **1510**. La endorribonucleasa unida puede escindir un módulo de ácido nucleico **1512** del agente de dirección genético multiplexado **1511** en una ubicación prescrita definida por la secuencia no nativa **1510**. La escisión **1525** puede procesar (por ejemplo, liberar) módulos de ácido nucleico individuales **1512**. En algunos aspectos, los módulos de ácido nucleico **1512** procesados pueden comprender la totalidad, una parte o nada de la secuencia no nativa **1510**. Los módulos de ácido nucleico 1512 procesados pueden estar unidos por un polipéptido dirigido a un sitio **1530**, formando de este modo un complejo **1531**. El complejo **1531** se puede dirigir **1535** a un ácido nucleico diana **1540**. El ácido nucleico diana **1540** puede ser escindido y/o modificado por el complejo **1531**.

Agentes de dirección genéticos multiplexados

Se puede usar un agente de dirección genético multiplexado para modificar múltiples ácidos nucleicos diana al mismo tiempo y/o en cantidades estequiométricas. Un agente de dirección genético multiplexado puede ser cualquier ácido nucleico dirigido a ácido nucleico como se describe en el presente documento en tándem. Un agente de dirección genético multiplexado puede referirse a una molécula continua de ácido nucleico que comprenda uno o más módulos de ácido nucleico. Un módulo de ácido nucleico puede comprender un ácido nucleico y una secuencia no nativa (por ejemplo, una fracción, secuencia de unión endorribonucleasa, ribozima). El ácido nucleico puede ser ARN no codificante tal como microARN (miARN), ARN de interferencia pequeño (ARNip), ARN largo no codificante (ARNlnc o ARNlinc), ARNip endógeno (endo-ARNip), ARN de interferente con piwi (ARNpiwi), ARN interferente corto de acción *trans* (ARNiptra), ARN interferente pequeño asociado a repetición (ARNipar), ARN nucleolar pequeño (ARNpno), ARN nuclear pequeño (ARNnp), ARN de transferencia (ARNt) y ARN ribosómico (ARNr), o cualquier combinación de los mismos. El ácido nucleico puede ser un ARN codificante (por ejemplo, un ARNm). El ácido nucleico puede ser cualquier tipo de ARN. En algunos aspectos, el ácido nucleico puede ser un ácido nucleico dirigido a ácido nucleico.

La secuencia no nativa puede ubicarse en el extremo 3’ del módulo de ácido nucleico. La secuencia no nativa puede ubicarse en el extremo 5' del módulo de ácido nucleico. La secuencia no nativa puede ubicarse tanto en el extremo 3’ como en el extremo 5' del módulo de ácido nucleico. La secuencia no nativa puede comprender una secuencia que puede unirse a una endorribonucleasa (por ejemplo, secuencia de unión a endorribonucleasa). La secuencia no nativa puede ser una secuencia reconocida con especificidad de secuencia por una endorribonucleasa (por ejemplo, RNasa T1 escinde bases de G no emparejadas, RNasa T2 escinde el extremo 3' de As, RNasa U2 escinde el extremo 3' de bases de A no emparejadas). La secuencia no nativa puede ser una secuencia que sea estructuralmente reconocida por una endorribonucleasa (por ejemplo, estructura de horquilla, uniones monocatenarias y bicatenarias, por ejemplo, Drosha reconoce una unión monocatenaria-bicatenaria dentro de una horquilla). La secuencia no nativa puede comprender una secuencia que puede unirse a una endorribonucleasa del sistema CRISPR (por ejemplo, proteína Csy4, Cas5 y/o Cas6). La secuencia no nativa puede comprender una secuencia de nucleótidos que tenga al menos o como máximo aproximadamente un 40 %, 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de identidad de secuencia y/o similitud de secuencia de nucleótidos con una de las siguientes secuencias:

5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3';
5'-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3';
5'-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3';
5'-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3';
5'-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3';
5'-GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3';
5'-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3';
5'-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3'.
5'-GUUGCAAGGGAUUGAGCCCCGUAAGGGGAUUGCGAC-3';
5'-GUUGCAAACCUCGUUAGCCUCGUAGAGGAUUGAAAC-3';
5'-GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC-3';
5'-GUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3';
5'-GAUAUAAACCUAAUUACCUCGAGAGGGGACGGAAAC-3';
5'-CCCCAGUCACCUCGGGAGGGGACGGAAAC-3';
5'-GUUCCAAUUAAUCUUAAACCCUAUUAGGGAUUGAAAC-3',
5'-GUCGCCCCCCACGCGGGGGCGUGGAUUGAAAC-3;
5'-CCAGCCGCCUUCGGGCGGCUGUGUGUUGAAAC-3';
5'-GUCGCACUCUACAUGAGUGCGUGGAUUGAAAU-3';
5'-UGUCGCACCUUAUAUAGGUGCGUGGAUUGAAAU-3 y
5'-GUCGCGCCCCGCAUGGGGCGCGUGGAUUGAAA-3'.

En algunos aspectos, en donde la secuencia no nativa comprende una secuencia de unión a endorribonucleasa, los módulos de ácido nucleico pueden estar unidos por la misma endorribonucleasa. Los módulos de ácido nucleico pueden no comprender la misma secuencia de unión a endorribonucleasa. Los módulos de ácido nucleico pueden comprender diferentes secuencias de unión a endorribonucleasa. Las diferentes secuencias de unión a endorribonucleasa pueden estar unidas por la misma endorribonucleasa. En algunos aspectos, los módulos de ácido nucleico pueden estar unidos por diferentes endorribonucleasas.

La fracción puede comprender una ribozima. La ribozima puede escindirse, liberando así cada módulo del agente de dirección genético multiplexado. Las ribozimas adecuadas pueden incluir ARNr de 23S de peptidil transferasa, RnasaP, intrones del grupo I, intrones del grupo II, ribozima de ramificación de GIR1, Leadzima, ribozimas de horquilla, ribozimas de cabeza de martillo, ribozimas de HDV, ribozimas de CPEB3, ribozimas de VS, ribozima de glmS, ribozima de CoTC y ribozimas sintéticas.

Los ácidos nucleicos de los módulos de ácido nucleico del agente de dirección genético multiplexado pueden ser idénticos. Los módulos de ácido nucleico pueden diferir en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos. Por ejemplo, los diferentes módulos de ácido nucleico pueden diferir en la región espaciadora del módulo de ácido nucleico, dirigiendo así el módulo de ácido nucleico a un ácido nucleico diana diferente. En algunos casos, los diferentes módulos de ácido nucleico pueden diferir en la región espaciadora del módulo de ácido nucleico, y aún así dirigirse al mismo ácido nucleico diana. Los módulos de ácido nucleico pueden dirigirse al mismo ácido nucleico diana. Los módulos de ácido nucleico pueden dirigirse a uno o más ácidos nucleicos diana.

Un módulo de ácido nucleico puede comprender una secuencia reguladora que puede permitir la traducción o amplificación apropiada del módulo de ácido nucleico. Por ejemplo, un módulo de ácido nucleico puede comprender un promotor, una caja de TATA, un elemento potenciador, un elemento de terminación de la transcripción, un sitio de unión al ribosoma, una región 3' sin traducir, una región 5' sin traducir, una secuencia de capuchón en 5', una secuencia de poliadenilación en 3', un elemento de estabilidad del ARN y similares.

Métodos

La divulgación proporciona métodos para la modificación de múltiples ácidos nucleicos diana, de forma simultánea, mediante el uso de un agente de dirección genético multiplexado. Un polipéptido dirigido a un sitio, una endorribonucleasa y un agente de dirección genético multiplexado se pueden introducir en una célula hospedadora. Se puede introducir un vector de la divulgación (por ejemplo, uno que comprende un agente de dirección genético multiplexado, una endorribonucleasa y/o un polipéptido dirigido a un sitio) en una célula hospedadora. En algunos casos, se puede introducir más de una endorribonucleasa y/o de un agente de dirección genético multiplexado en las células. Si un agente de dirección genético multiplexado comprende diferentes tipos de fracciones, donde los fracciones son diferentes secuencias de unión a endorribonucleasa, entonces una o más endorribonucleasas correspondientes a los tipos de secuencias de unión del agente de dirección genético multiplexado pueden introducirse en las células.

La introducción se puede producir mediante cualquier medio de introducción de un ácido nucleico en una célula, tal como infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares. El vector puede

expresarse transitoriamente en la célula hospedadora. El vector puede expresarse de manera estable en la célula hospedadora (por ejemplo, mediante la integración estable en el genoma de la célula hospedadora).

En los casos en los que un fracción comprende una secuencia de unión a endorribonucleasa, se puede expresar una endorribonucleasa y puede unirse a un sitio de unión a endorribonucleasa en el agente de dirección genético multiplexado. La endorribonucleasa puede escindir el agente de dirección genético multiplexado en módulos de ácido nucleico individuales.

En los casos en los que un fracción comprende una ribozima, es posible que no se requiera la expresión de una endorribonucleasa en una célula hospedadora. La ribozima puede escindirse, lo que produce la escisión del agente de dirección genético multiplexado en módulos de ácido nucleico individuales.

Los módulos de ácido nucleico individuales (por ejemplo, escindidos) pueden comprender toda, una parte o nada de la fracción (por ejemplo, secuencia de unión a endorribonucleasa). Por ejemplo, el módulo de ácido nucleico liberado (por ejemplo, procesado) puede someterse a un recorte y/o una degradación por exonucleasa que puede producir la eliminación del extremo 5' y/o 3' del módulo de ácido nucleico. En dichos casos, el recorte y/o la degradación por exonucleasa puede producir la eliminación de toda, parte o nada de la fracción (por ejemplo, secuencia de unión a endorribonucleasa).

El módulo de ácido nucleico liberado (por ejemplo, procesado) puede unirse a un polipéptido dirigido a un sitio, formando así un complejo. El complejo puede ser guiado a un ácido nucleico diana por el ácido nucleico dirigido a ácido nucleico que puede hibridarse con el ácido nucleico diana de una manera específica de la secuencia. Una vez hibridado, el polipéptido dirigido a un sitio del complejo puede modificar el ácido nucleico diana (por ejemplo, escindir el ácido nucleico diana). En algunos casos, la modificación comprende la introducción de una rotura bicatenaria en el ácido nucleico diana. En algunos casos, la modificación comprende la introducción de una rotura monocatenaria en el ácido nucleico diana.

En algunos aspectos, uno o más polinucleótidos y/o vectores donantes que codifican los mismos pueden introducirse en la célula. Se pueden incorporar uno o más polinucleótidos donantes en los ácidos nucleicos diana modificados (por ejemplo, escindidos), lo que produce una inserción. El mismo polinucleótido donante puede incorporarse en múltiples sitios de escisión de ácidos nucleicos diana. Se pueden incorporar uno o más polinucleótidos donantes en uno o más sitios de escisión de ácidos nucleicos diana. Esto se puede denominar ingeniería de genoma multiplex. En algunos casos, no se puede introducir en las células ningún polinucleótido ni/o vector donante que lo codifique. En estos casos, el ácido nucleico diana modificado puede comprender una eliminación.

Administración estequiométrica de ácidos nucleicos

Visión general

La divulgación proporciona composiciones, métodos y kits para la administración estequiométrica de un ácido nucleico a una ubicación celular y/o subcelular. La administración estequiométrica puede estar mediada por un complejo. La **Figura 16** representa un complejo ilustrativo para la administración estequiométrica de una pluralidad de ácidos nucleicos a una célula y/o ubicación subcelular. El complejo puede comprender una pluralidad de ácidos nucleicos **1605**. Cada ácido nucleico puede comprender un sitio de unión de proteínas de unión a ácido nucleico **1610**. Los sitios de unión de proteínas de unión a ácido nucleico **1610** pueden ser todas las mismas secuencias, secuencias diferentes, o algunas pueden ser las mismas secuencias y algunas pueden ser secuencias diferentes. En algunos aspectos, los sitios de unión de proteínas de unión a ácido nucleico pueden unirse a un miembro de la familia Cas6, Cas5 o Csy4. El complejo puede comprender un polipéptido de fusión en tándem **1630**. El polipéptido de fusión en tándem puede comprender proteínas de unión a ácido nucleico **1625** fusionadas entre sí en tándem. Las proteínas de unión a ácido nucleico pueden estar separadas por un enlazador **1620**. Las proteínas de unión a ácido nucleico **1625** pueden ser la misma proteína, pueden ser proteínas diferentes, o algunas pueden ser las mismas proteínas y algunas pueden ser proteínas diferentes. Las proteínas de unión a ácido nucleico **1625** pueden ser proteínas Csy4. Las proteínas de unión a ácido nucleico **1625** pueden unirse a un sitio de unión de proteína de unión a ácido nucleico **1610** sobre el ácido nucleico **1605**. El polipéptido de fusión en tándem **1630** puede comprender una secuencia no nativa **1615**. En algunos casos, la secuencia no nativa es una secuencia de ubicación subcelular (por ejemplo, nuclear). En algunos aspectos, el ácido nucleico **1605** puede codificar una secuencia no nativa (por ejemplo, una secuencia de ubicación subcelular (por ejemplo, nuclear)). El complejo se puede introducir **1635** en las células, en donde uno o más de los ácidos nucleicos **105** se pueden traducir en los polipéptidos **1640**. Un polipéptido traducido **1640** puede unirse y escindir el sitio de unión de la proteína de unión a ácido nucleico **1610** en el ácido nucleico **1605**. La escisión **1645** puede liberar el ácido nucleico **1650** que puede ser un ácido nucleico dirigido a ácido nucleico. El ácido nucleico liberado **1650** puede unirse a un polipéptido traducido **1645** (por ejemplo, un polipéptido dirigido a un sitio), formando así una unidad. El polipéptido traducido **1645** puede comprender una señal de localización nuclear. La unidad puede translocarse al núcleo, en donde la unidad puede guiarse a un ácido nucleico diana hibridable con el espaciador del ácido nucleico liberado **1650**. La unidad puede hibridarse a un ácido nucleico diana. El polipéptido dirigido a un sitio de la unidad puede escindir el ácido nucleico diana. La escisión del ácido nucleico diana puede denominarse ingeniería genómica. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a

ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

En algunos aspectos, los múltiples ácidos nucleicos dirigidos a ácido nucleico pueden administrarse estequiométricamente en una ubicación celular y/o subcelular. La **Figura 17** representa un complejo ilustrativo para la administración estequiométrica de una pluralidad de ácidos nucleicos. El complejo puede comprender una pluralidad de ácidos nucleicos **1705**. Cada ácido nucleico puede comprender una pluralidad de sitios de unión de proteínas de unión a ácido nucleico **1710**/**1711**. Los sitios de unión de proteínas de unión a ácido nucleico **1710/1711** pueden ser todas las mismas secuencias, secuencias diferentes, o algunas pueden ser las mismas secuencias y algunas pueden ser secuencias diferentes. En algunos aspectos, los sitios de unión de proteínas de unión a ácido nucleico **1710**/**1711** pueden unirse a un miembro de la familia Cas6, Cas5 o Csy4. El complejo puede comprender un polipéptido de fusión en tándem **1730**. El polipéptido de fusión en tándem puede comprender proteínas de unión a ácido nucleico **1725** fusionadas entre sí en tándem. Las proteínas de unión a ácido nucleico pueden estar separadas por un enlazador **1720**. Las proteínas de unión a ácido nucleico **1725** pueden ser la misma proteína, pueden ser proteínas diferentes, o algunas pueden ser las mismas proteínas y algunas pueden ser proteínas diferentes. Las proteínas de unión a ARN **1725** pueden ser una combinación de polipéptidos Csy4, Cas5 y Cas6. Las proteínas de unión a ácido nucleico **1725** pueden unirse a un sitio de unión de proteína de unión a ácido nucleico **1710** sobre el ácido nucleico **1705**. El polipéptido de fusión en tándem **1730** puede comprender una secuencia no nativa **1715**. En algunos casos, la secuencia no nativa es una secuencia de ubicación subcelular (por ejemplo, nuclear). En algunos aspectos, el ácido nucleico **1705** puede codificar una secuencia no nativa (por ejemplo, una secuencia de ubicación subcelular (por ejemplo, nuclear)). El complejo puede introducirse en las células **1735**, en donde uno o más de los ácidos nucleicos pueden traducirse en los polipéptidos **1740**/**1750**. Un polipéptido traducido **1740** puede unirse y escindir el sitio de unión de la proteína de unión a ácido nucleico **1711** en el ácido nucleico **1705**. La escisión **1745** puede liberar el ácido nucleico **1755**, que puede ser un ácido nucleico dirigido a ácido nucleico y/o un polinucleótido donante. El ácido nucleico liberado **1755** puede unirse a un polipéptido traducido **1750** (por ejemplo, un polipéptido dirigido a un sitio), formando así una unidad. En algunos casos, El polipéptido traducido **1750** comprende una señal de localización nuclear. La unidad puede translocarse al núcleo, en donde la unidad puede guiarse a un ácido nucleico diana hibridable con el espaciador del ARN liberado **1755**. La unidad puede hibridarse a un ácido nucleico diana. El polipéptido dirigido a un sitio de la unidad puede escindir el ácido nucleico diana.

Métodos

La divulgación proporciona métodos para la administración estequiométrica de ácidos nucleicos a una célula (por ejemplo, ácidos nucleicos administrables estequiométricamente). El método puede comprender la unión de un polipéptido de fusión en tándem a una pluralidad de ácidos nucleicos administrables estequiométricamente, formando así un complejo. El complejo puede comprender cantidades estequiométricas de los ácidos nucleicos (por ejemplo, el complejo puede comprender la pluralidad de ácidos nucleicos en una proporción y/o cantidad prescrita). Se pueden administrar 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más ácidos nucleicos estequiométricamente. En algunos casos, se pueden administrar 3 ácidos nucleicos administrables estequiométricamente. En algunos casos, se pueden administrar 4 ácidos nucleicos administrables estequiométricamente.

El ácido nucleico administrable estequiométricamente puede codificar un polipéptido o un ARN no codificante. El polipéptido puede ser un polipéptido del sistema CRISPR (por ejemplo, un polipéptido dirigido a un sitio, una endorribonucleasa). El ácido nucleico administrable estequiométricamente puede codificar más de un polipéptido. El ácido nucleico administrable estequiométricamente puede comprender una pluralidad de ácidos nucleicos administrables estequiométricamente (por ejemplo, en una matriz). El ácido nucleico administrable estequiométricamente puede codificar un ARN no codificante. Los ejemplos de ARN no codificantes pueden incluir microARN (miARN), ARN de interferencia pequeño (ARNip), ARN largo no codificante (ARNlnc o ARNlinc), ARNip endógeno (endo-ARNip), ARN de interferente con piwi (ARNpiwi), ARN interferente corto de acción *trans* (ARNiptra), ARN interferente pequeño asociado a repetición (ARNipar), ARN nucleolar pequeño (ARNpno), ARN nuclear pequeño (ARNnp), ARN de transferencia (ARNt) y ARN ribosómico (ARNr). El ácido nucleico administrable estequiométricamente puede ser ARN.

El ácido nucleico administrable estequiométricamente puede codificar una secuencia no nativa. En algunos casos, el ácido nucleico administrable estequiométricamente codifica una secuencia no nativa de modo que cuando un polipéptido se traduce de un ácido nucleico administrable estequiométricamente que codifica un polipéptido, el polipéptido se fusiona con la secuencia no nativa (por ejemplo, generando una proteína de fusión). La secuencia no nativa puede ser un marcador de afinidad peptídico. La secuencia no nativa (por ejemplo, marcador de afinidad peptídico) se puede ubicar en el extremo N del polipéptido, el extremo C del polipéptido o cualquier ubicación dentro del polipéptido (por ejemplo, un lazo accesible desde la superficie). En algunos aspectos, la secuencia no nativa es una señal de localización nuclear (NLS). Una NLS puede ser secuencia monopartita o bipartita. La NLS puede ser reconocida por la maquinaria de importación nuclear (por ejemplo, las importinas). Una NLS puede ser un péptido pequeño (por ejemplo, PKKKRKV del antígeno t grande del SV40). Una NLS puede ser un dominio polipeptídico (por ejemplo, dominio M9 ácido de hnRNP A1).

La secuencia no nativa puede ser un marcador de afinidad de ácido nucleico (por ejemplo, señal de localización de

ácido nucleico). Por ejemplo, un ácido nucleico administrable estequiométricamente que codifica un ADN (por ejemplo, un polinucleótido donante) puede comprender una señal de localización de ácido nucleico que puede localizar el ADN en el núcleo. Dichas señales de localización de ácido nucleico pueden incluir, por ejemplo, secuencias de ácido nucleico peptídico (PNA).

Los ácidos nucleicos administrables estequiométricamente pueden comprender secuencias reguladoras que pueden permitir la traducción o amplificación apropiada del ácido nucleico. Por ejemplo, un ácido nucleico puede comprender un promotor, una caja de TATA, un elemento potenciador, un elemento de terminación de la transcripción, un elemento de estabilidad del ADN, un sitio de unión al ribosoma, una región 3' sin traducir, una región 5' sin traducir, una secuencia de capuchón en 5', una secuencia de poliadenilación en 3', un elemento de estabilidad del ARN y similares.

El ácido nucleico puede comprender un sitio de unión de proteínas de unión a ácido nucleico. El sitio de unión de proteína de unión a ácido nucleico puede estar unido por un una proteína de unión a ácido nucleico. El sitio de unión de proteína de unión a ácido nucleico puede estar unido por un polipéptido CRISPR (por ejemplo, polipéptido dirigido a un sitio, una endorribonucleasa). El sitio de unión de proteína de unión a ácido nucleico puede estar unido por un polipéptido de la familia Cas5 o Cas6. El sitio de unión de proteína de unión a ácido nucleico puede estar unido por un polipéptido Csy4, Cas5 o Cas6. Algunos ejemplos de sitios de unión de proteínas de unión a ácido nucleico pueden incluir, por ejemplo, secuencias que pueden ser unidas por proteínas de unión a ARN tales como la secuencia de unión a MS2, la secuencia de unión a U1A, la secuencia de boxB, la secuencia de eIF4A, horquillas, secuencias que pueden ser unidas por dominios de motivos de reconocimiento de ARN (RRM) (por ejemplo, U1A), secuencias que pueden estar unidas por dominios de unión a ARN bicatenarios (dsRBD) (por ejemplo, Staufen), secuencias que pueden estar unidas por dominios PAZ (por ejemplo, PAZ, Argonauta), secuencias que pueden estar unidas por dominios PIWI (por ejemplo, PIWI, MILI, MIWI, Argonauta) y similares,. Algunos ejemplos de sitios de unión de proteínas de unión a ácido nucleico pueden incluir, por ejemplo, secuencias que pueden estar unidas por proteínas de unión a ADN tales como dedos de cinc, un dominio hélice-vuelta-hélice, un dominio de dedo de cinc, un dominio de cremallera de leucina (bZIP), un dominio de hélice alada, un dominio de hélice alada-vuelta-hélice, un dominio hélice-lazo-hélice, un dominio de caja HMG, un dominio Wor3, un dominio de inmunoglobulina, un dominio B3, un dominio TALE y similares.

El ácido nucleico puede comprender uno o más sitios de unión de proteínas de unión a ácido nucleico. El ácido nucleico puede comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más sitios de unión de proteínas de unión a ácido nucleico. El uno o más sitios de unión de proteínas de unión a ácido nucleico pueden ser iguales. El uno o más sitios de unión de proteínas de unión a ácido nucleico pueden ser diferentes. Por ejemplo, el ácido nucleico puede comprender un sitio de unión a Csy4 y un sitio de unión a MS2. El uno o más sitios de unión de proteínas de unión a ácido nucleico pueden estar separados por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500 o más nucleótidos. En algunos aspectos, el sitio de unión de la proteína de unión al ácido nucleico más a 3’ puede estar unido por un polipéptido de fusión en tándem de la divulgación.

Polipéptido de fusión en tándem

En algunos aspectos, el método de la divulgación proporciona la unión de una pluralidad de ácidos nucleicos a un polipéptido de fusión en tándem. Un polipéptido de fusión en tándem puede comprender una pluralidad de proteínas de unión a ácido nucleico fusionadas en una cadena polipeptídica. Un polipéptido de fusión en tándem puede comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más proteínas de unión a ácido nucleico. Las proteínas de unión a ácido nucleico del polipéptido de fusión en tándem pueden unirse a los sitios de unión de proteína de unión a ácido nucleico de los ácidos nucleicos de la divulgación. Los ejemplos de proteínas de unión a ácido nucleico pueden incluir MS2, U1A, proteínas de unión a secuencia de boxB (por ejemplo, dedos de cinc), eIF4A, Staufen, PAZ, Argonauta, PIWI, MILI, MIWI, dedos de cinc, un dominio hélice-vuelta-hélice, un dominio de dedo de cinc, un dominio de cremallera de leucina (bZIP), un dominio de hélice alada, un dominio de hélice alada-vuelta-hélice, un dominio hélice-lazo-hélice, un dominio de caja HMG, un dominio Wor3, un dominio de inmunoglobulina, un dominio B3, un dominio TALE y similares. En algunos aspectos, la proteína de unión a ácido nucleico es una proteína de unión a ARN. La proteína de unión a ARN puede ser miembro de un sistema CRISPR. En algunos aspectos, la proteína de unión a ARN puede ser un miembro de la familia de proteínas Cas5 o Cas6. En algunos aspectos, la proteína de unión a ARN puede ser Csy4, Cas5, Cas6 o cualquier combinación de las mismas. En algunos aspectos, la proteína de unión a ácido nucleico es una proteína de unión a ADN (por ejemplo, un dedo de cinc).

En algunos casos, Las proteínas de unión a ácido nucleico están separadas por un enlazador. Un enlazador puede comprender aproximadamente 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más aminoácidos.

Un polipéptido de fusión en tándem puede comprender una secuencia no nativa (por ejemplo, marcador de afinidad peptídico). La secuencia no nativa puede comprender una señal de localización nuclear (NLS) que puede dirigir el polipéptido de fusión en tándem a una ubicación subcelular (por ejemplo, núcleo).

Cada proteína de unión a ácido nucleico del polipéptido de fusión en tándem puede comprender su propia secuencia no nativa. La secuencia no nativa de cada proteína de unión a ácido nucleico puede ser la misma. La secuencia no nativa de cada proteína de unión a ácido nucleico puede ser diferente. La secuencia no nativa de algunas de las

proteínas de unión a ácido nucleico del polipéptido de fusión en tándem puede ser la misma y la secuencia no nativa de algunas de las proteínas de unión a ácido nucleico del polipéptido de fusión en tándem puede ser diferente.

En algunos casos, los métodos de la divulgación pueden proporcionar la formación de un complejo que comprenda un polipéptido de fusión en tándem y una pluralidad de ácidos nucleicos de la divulgación. La formación del complejo puede comprender la unión de las proteínas de unión a ácido nucleico del polipéptido de fusión en tándem a su secuencia de unión a proteína de unión a ácido nucleico afín en los ácidos nucleicos de la divulgación. Por ejemplo, un ácido nucleico administrable estequiométricamente que comprende un sitio de unión a Csy4, puede unirse a la subunidad de proteína Csy4 en la proteína de fusión en tándem. El complejo puede formarse fuera de las células (por ejemplo, *in vitro*). El complejo puede formarse en células (por ejemplo, *in vivo*). Cuando se forma un complejo *in vitro*,se puede introducir en una célula mediante, por ejemplo, transfección, transformación, transducción vírica, electroporación, inyección y similares.

Los métodos de la divulgación proporcionan la administración terapéutica de múltiples ácidos nucleicos *in vivo*, *in vitro* y *ex vivo*. Los ácidos nucleicos administrados pueden usarse para tratar una enfermedad. Por ejemplo, los ácidos nucleicos administrados pueden usarse en terapia génica y/o pueden integrarse en el genoma de la célula, proporcionando así un resultado terapéutico. Un resultado terapéutico puede referirse al aumento o a la disminución de los niveles de una proteína, ácido nucleico o cualquier molécula biológica relacionada con una enfermedad, tal como un producto de degradación, una molécula pequeña y/o un ion. Por ejemplo, un resultado terapéutico puede comprender aumentar los niveles de un gen antiinflamatorio o disminuir los niveles de una proteína en una vía relacionada con una enfermedad. Un resultado terapéutico puede referirse a un efecto fisiológico. Los efectos fisiológicos pueden incluir cambios morfológicos, cambios metabólicos y/o cambios estructurales en una célula. Un resultado terapéutico puede referirse a cambios en las modificaciones de una proteína y/o un ácido nucleico, tales como glicosilación, acetilación, metilación, desmetilación, despurinación, ubiquitinilación y similares.

Un resultado terapéutico puede medirse mediante cambios en la composición genética de la célula, los niveles de biomoléculas de interés en la célula y/o los cambios fisiológicos en la célula. Las mediciones se pueden realizar usando técnicas de biología molecular, tales como espectroscopia, espectrometría, secuenciación, ELISA, microscopía y/o cristalografía de rayos X. Las mediciones se pueden realizar usando modelos animales, tales como ratón, ratas, perros y primates. Por ejemplo, las células modificadas genéticamente de la divulgación pueden introducirse en ratones y evaluarse para detectar cambios biológicos y fisiológicos tales como, por ejemplo, la capacidad de hacer metástasis y/o diferenciarse.

Espaciadores para trastornos de la sangre

La presente divulgación proporciona composiciones, métodos y kits para la ingeniería genética de células madre hematopoyéticas (HSC).

Composiciones

Una HSC puede comprender un polipéptido dirigido a un sitio (por ejemplo, Cas9).

Una HSC puede comprender un ácido nucleico dirigido a ácido nucleico. El ácido nucleico dirigido a ácido nucleico de la divulgación puede dirigirse a un gen implicado en un trastorno genético. La Tabla 1 enumera genes ilustrativos que participan en trastornos. Los genes enumerados en la Tabla 1 pueden ser genes que pueden ser la diana de un ácido nucleico dirigido a ácido nucleico. El ácido nucleico dirigido a ácido nucleico de la divulgación puede comprender un espaciador que puede dirigirse a un gen enumerado en la Tabla 2.

La Tabla 2 representa separadores ilustrativos de los ácidos nucleicos dirigidos a ácidos nucleicos de la divulgación. Cada espaciador de la Tabla 2 puede ser un espaciador que se puede insertar en un ácido nucleico dirigido a ácido nucleico. Los espaciadores ilustrativos están acompañados por el nombre del trastorno y el gen que participa en el trastorno que es la diana de los espaciadores.

Tabla 2. Espaciadores de los trastornos sanguíneos.

Gen	Espaciador	Enfermedad
Arilsulfatasa A	AGGGTTTATTTTCTTACGCT	Leucodistrofia metacromática
Proteína del síndrome de Wiskott-Aldrich	CCGAGGTCCCTAGTCCGGAA	Síndrome de Wiskott-Aldrich
Casete de unión a ATP D1	CGGAGGTGGGCGGAGCCTCC	Adreno-leucodistrofia
Receptor de quimiocina C-C de tipo 5	AGCTTTCTCGTCTGGGTATT	Virus de la inmunodeficiencia humana

(continuación)

Gen	Espaciador	Enfermedad
Subunidad beta de hemoglobina	GAAAATAAATGTTTTTTATT	Beta talasemia
Subunidad gamma del receptor de interleucina-2	ACAGAAACTTTATTTCTCAT	Identificación combinada severa ligada a X
Cistinosina	CTTTGGGAGGCCGAGGCGGG	Cistinosis MLSD
Proteína ribosómica S19	TTTTAGAAACAGTATGAGAT	Anemia de Diamon-Blackfan
Complemento de anemia de Fanconi grupo B	ATGCACAAAATAAACAGCAG	Anemia de Fanconi
Gen del síndrome de Shwachman-Bodian-Diamond	GAGTTAGTTCACATCTACAG	Síndrome de Shwachman-Bodian-Diamond

Métodos

La divulgación proporciona métodos para introducir un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio en una HSC. En algunos aspectos, la HSC se extrae de un paciente antes de la introducción. La HSC extraída se puede purificar (por ejemplo, por aféresis). El polipéptido dirigido a un sitio y/o el ácido nucleico dirigido a ácido nucleico pueden ser introducidos en una HSC que haya sido purificada. El polipéptido dirigido a un sitio y/o el ácido nucleico dirigido a ácido nucleico se pueden introducir en una HSC que no se haya purificado. La introducción puede ocurrir en una HSC *in vitro* (por ejemplo, fuera del paciente, célula extraída). En algunos casos, la introducción ocurre en una HSC *in vivo* (por ejemplo, dentro de un paciente, célula no extraída).

La introducción del polipéptido dirigido a un sitio y/o del ácido nucleico dirigido a ácido nucleico de la divulgación puede ocurrir, por ejemplo, transducción vírica, transfección, electroporación, transfección óptica y/o transfección química.

Una vez introducido en una HSC, el ácido nucleico dirigido a ácido nucleico de la divulgación y el polipéptido dirigido a un sitio pueden formar un complejo. El complejo puede ser guiado a un ácido nucleico diana (por ejemplo, los genes enumerados en la Tabla 3) por el ácido nucleico dirigido a ácido nucleico. El ácido nucleico dirigido a ácido nucleico puede hibridarse con el ácido nucleico diana. El polipéptido dirigido a un sitio puede modificar el ácido nucleico diana (por ejemplo, mediante la escisión del ácido nucleico diana).

En algunos casos, el ácido nucleico diana modificado comprende una eliminación. En algunos casos, el ácido nucleico diana modificado comprende una eliminación comprende una inserción de un polinucleótido donante. Un polinucleótido donante, una parte de un polinucleótido donante, una copia de un polinucleótido donante o una parte de una copia de un polinucleótido donante en un ácido nucleico diana. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

Tabla 3. Lista de genes que participan en enfermedades

Nombre de la enfermedad	**Gen/es**
Leucodistrofia metacromática (MLD)	Arilsulfatasa A
Síndrome de Wiskott-Aldrich (WAS)	Proteína del síndrome de Wiskott-Aldrich
Síndrome de Wiskott-Aldrich (WAS)	Leucosialina
Neutropenia	Proteína del síndrome de Wiskott-Aldrich
Adreno-leucodistrofia	Casete de unión a ATP D1
Virus de la inmunodeficiencia humana (VIH)	Receptor de quimiocina C-C de tipo 5
Beta talasemia	subunidad beta de hemoglobina
Anemia falciforme	subunidad beta de hemoglobina
Inmunodeficiencia combinada severa ligada al cromosoma X (X-SCID)	Subunidad gamma del receptor de interleucina-2
Trastorno de almacenamiento lisosómico multisistémico cistinosis	Cistinosina
Anemia de Diamond-Blackfan	Proteína ribosómica S19
Anemia de Fanconi	Grupos de complementación de anemia de Fanconi A, B y C
Síndrome de Shwachman-Bodian-Diamond	Gen SBDS
Enfermedad de Gaucher	Glucocerebrosidasa

(continuación)

Nombre de la enfermedad	Gen/es
Hemofilia A	Factor antihemofílico O Factor VIII
Hemofilia B	Factor de Christmas, Serina proteasa, Factor IX
Deficiencia de adenosina desaminasa (ADA-SCID)	Adenosina desaminasa
Gangliosidosis GM1	beta-galactosideasa
Enfermedad del almacenamiento del glucógeno de Tipo II, Enfermedad de Pompe, deficiencia de maltasa ácida	alfa-glucosidasa ácida
enfermedad de Niemann-Pick, SMPD1 asociado (Tipos A y B)	Esfingomielina fosfodiesterasa 1 O esfingomielinasa ácida
enfermedad de Krabbe, leucodistrofia de células globoides, lipidosis de galactosilceramida	Galactosilceramidasa O galactercerebrosideasa
Esclerosis múltiple (EM)	Antígenos leucocitarios humanos DR-15, DQ-6, DRB1

Métodos computacionales

La divulgación proporciona métodos computacionales para identificar espaciadores para ácidos nucleicos dirigidos a ácido nucleico. El método computacional puede comprender el barrido de la secuencia de ácido nucleico de un genoma en busca de un motivo adyacente al protoespaciador. Al encontrar un motivo adyacente al protoespaciador, el programa puede contar automáticamente entre 10-30 nucleótidos cadena arriba del motivo adyacente al protoespaciador. Los 10-30 nucleótidos cadena arriba del motivo adyacente al protoespaciador pueden constituir una secuencia espaciadora putativa. En otras palabras, los 10-30 nucleótidos cadena arriba del motivo adyacente al protoespaciador en el genoma pueden corresponder a un ácido nucleico diana, y una secuencia complementaria al ácido nucleico diana puede denominarse espaciador.

El programa puede probar cada iteración de secuencia de la secuencia espaciadora putativa para determinar la eficacia de la secuencia como espaciador en un ácido nucleico dirigido a ácido nucleico. Por ejemplo, el programa puede tomar cada iteración de la secuencia espaciadora putativa y realizar una predicción de estructura secundaria *in silico* en la secuencia. La predicción de la estructura secundaria puede comprender agregar las secuencias espaciadoras putativas a una cadena principal de ácido nucleico dirigido a ácido nucleico (por ejemplo, el ácido nucleico dirigido a ácido nucleico sin el espaciador). La predicción de la estructura secundaria puede realizar un análisis de predicción de la estructura secundaria de la secuencia espaciadora putativa acoplada en la cadena principal de ácido nucleico dirigido a ácido nucleico. El análisis de predicción de la estructura secundaria puede comprender, por ejemplo, predecir qué nucleótidos pueden formar dúplex, horquillas, qué nucleótidos no están estructurados y/o qué nucleótidos pueden estar desapareados.

El método computacional puede comprender implementar una prueba de plegamiento en cada secuencia espaciadora putativa que haya sido sometida a un análisis de predicción de la estructura secundaria. La prueba de plegamiento puede comprender el plegamiento *in silico* del ácido nucleico dirigido a ácido nucleico que comprende la secuencia espaciadora putativa. El ácido nucleico dirigido a ácido nucleico y la secuencia espaciadora putativa pueden pasar o no la prueba de plegamiento.

Para pasar la prueba de plegamiento, la estructura secundaria de la cadena principal de ácido nucleico dirigido a ácido nucleico puede necesitar conservarse, menos de 5, 4, 3, 2 o 1 nucleótidos en el espaciador se hibridan con nucleótidos fuera del espaciador y otra segunda estructura del espaciador está contenida dentro del espaciador.

Selección perfecta del indicador

Visión general

La presente divulgación describe métodos, composiciones, sistemas y kits para la modificación genética de células y la selección de dichas células modificadas genéticamente mediante la incorporación, detección y escisión perfectas de un elemento indicador. En algunos aspectos de la divulgación, un polinucleótido donante puede comprender un ácido nucleico que se va a introducir en el genoma de una célula (en el presente documento denominado elemento genético de interés), así como una secuencia de ácido nucleico que codifica un elemento indicador (por ejemplo, GFP), un polipéptido dirigido a un sitio y dos ácidos nucleicos dirigidos a ácido nucleico. El polipéptido dirigido a un sitio, los ácidos nucleicos dirigidos a ácido nucleico y/o los tres pueden ser controlados por un promotor inducible. Un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico pueden formar un complejo que puede dirigirse a un sitio en el genoma celular por hibridación del ácido nucleico dirigido a ácido nucleico con un ácido nucleico diana en el genoma. El polipéptido dirigido a un sitio del complejo puede escindir el ácido nucleico diana. El polinucleótido donante puede insertarse en el ácido nucleico diana escindido. Tras la introducción de una rotura bicatenaria (o rotura monocatenaria) en el sitio diana en presencia del polinucleótido donante, la población de células receptoras puede cribarse para detectar la presencia de la molécula indicadora como un indicador de la presencia del

elemento genético de interés. Tras el aislamiento de las células que contienen moléculas indicadoras, el elemento indicador se puede eliminar mediante la inducción de la expresión del polipéptido dirigido a un sitio y/o de ácido nucleico dirigido a ácido nucleico. Los ácidos nucleicos dirigios a ácido nucleico pueden dirigirse a los extremos 5' y 3' del elemento indicador y pueden provocar la escisión del elemento indicador. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

La **FIGURA 18** representa un método ilustrativo de la divulgación. Un ácido nucleico puede comprender una pluralidad de elementos genéticos **1805**/**1810**. Los elementos genéticos **1805** y **1810** pueden ser, por ejemplo, genes, ácidos nucleicos no codificantes, intrones, exones, ADN y/o ARN. Los elementos genéticos **1805** y **1810** pueden ser partes del mismo gen. Entre los elementos genéticos puede haber un ácido nucleico diana **1806** adecuado para la ingeniería genética. Un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico de la divulgación pueden formar un complejo que puede dirigirse a **1815** al ácido nucleico diana **1806**. El polipéptido dirigido a un sitio del complejo puede escindir **1820** el ácido nucleico diana **1806**. Un polinucleótido donante puede insertarse **1825** en el ácido nucleico diana escindido **1806**. El polinucleótido donante puede comprender un elemento genético de interés **1830**. El elemento genético de interés **1830** puede ser un gen. El polinucleótido donante también puede comprender un elemento indicador **1835**. El polinucleótido donante también puede comprender una secuencia polinucleotídica que codifica un polipéptido dirigido a un sitio y uno o más ácidos nucleicos dirigidos a ácido nucleico. En algunos casos, el polinucleótido que codifica el polipéptido dirigido a un sitio y el uno o más ácidos nucleicos dirigidos a ácido nucleico codifica dos ácidos nucleicos dirigidos a ácido nucleico. La secuencia polinucleotídica que codifica el polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico puede unirse operativamente a un promotor inducible. La inserción del polinucleótido donante en el ácido nucleico diana **1806** puede producir la expresión del elemento indicador **1835**. El elemento indicador **1835** puede usarse como una forma de seleccionar células que comprenden el polinucleótido donante.

La **FIGURA 19** representa la eliminación del elemento indicador **1915** del ácido nucleico diana. Un ácido nucleico diana puede comprender una pluralidad de elementos genéticos **1905**/**1920**. El elemento indicador **1915** puede fusionarse con un elemento genético de interés **1910**. La expresión del gen indicador **1915** puede inducirse, lo que puede producir la producción de un polipéptido dirigido a un sitio y uno o más ácidos nucleicos dirigidos a ácido nucleico. El polipéptido dirigido a un sitio puede formar complejos con los ácidos nucleicos dirigidos a ácido nucleico. Los complejos pueden ser guiados al elemento indicador **1915** por el ácido nucleico dirigido a ácido nucleico del complejo. Uno de los dos ácidos nucleicos dirigidos a ácido nucleico puede dirigirse **1925** al extremo 5' del elemento indicador **1915**. Uno de los dos ácidos nucleicos dirigidos a ácido nucleico pueden dirigirse **1930** al extremo 3' del elemento indicador **1915**. Los extremos diana del elemento indicador **1915** pueden ser escindidos por el polipéptido dirigido a un sitio del complejo, eliminando **1935** así el elemento indicador **1915**. El ácido nucleico diana puede comprender la parte del elemento genético de interés **1910** del polinucleótido donante. Los ácidos nucleicos dirigidos a ácido nucleico pueden diseñarse de modo que se escinda el polinucleótido donante (incluyendo el elemento genético de interés).

Métodos

La presente divulgación proporciona métodos para seleccionar células usando un elemento indicador y la escisión del elemento indicador. Un polipéptido dirigido a un sitio, una endorribonucleasa, un ácido nucleico dirigido a ácido nucleico, un polinucleótido donante y/o un ácido nucleico dirigido a un ácido nucleico pueden introducirse en una célula. El polinucleótido donante puede incluir uno o más elementos genéticos de interés. El polinucleótido donante puede incluir uno o más elementos indicadores. El polinucleótido donante incluye uno o más elementos genéticos de interés y uno o más elementos indicadores. Se puede introducir en una célula más de un polipéptido dirigido a un sitio, endorribonucleasa, polinucleótido donante y/o ácido nucleico dirigido a ácido nucleico. En algunos casos, la célula expresa ya un polipéptido dirigido a un sitio y/o un ácido nucleico dirigido a ácido nucleico. En algunos casos, el polipéptido dirigido a un sitio y/o el ácido nucleico dirigido a ácido nucleico son codificados en un plásmido. En algunos casos, el polipéptido dirigido a un sitio y/o el ácido nucleico dirigido a ácido nucleico está codificado en más de un plásmido. En algunos casos, se introduce en la célula más de un polipéptido dirigido a un sitio o ácido nucleico dirigido que codifica un polipéptido dirigido a un sitio. En algunos casos, la célula es un lisado celular.

La introducción se puede producir mediante cualquier medio de introducción de un ácido nucleico en una célula, tal como infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, transfección con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares. El vector puede expresarse transitoriamente en la célula hospedadora. El vector puede expresarse de manera estable en la célula hospedadora (por ejemplo, mediante la integración estable en el genoma de la célula hospedadora).

Un ácido nucleico dirigido a ácido nucleico puede unirse a un ácido nucleico caracterizado por una secuencia diana particular y/o cualquier secuencia homóloga a una secuencia particular. La secuencia diana puede ser parte o la totalidad de un gen, un extremo 5' de un gen, un extremo 3' de un gen, un elemento regulador (por ejemplo, promotor,

potenciador), un pseudogén, ADN no codificante, un microsatélite, un intrón, un exón, ADN cromosómico, ADN mitrocondrial, ADN sensible, ADN no codificante, ADN nucleoide, ADN o ARN de cloroplasto entre otras entidades de ácido nucleico.

El polipéptido dirigido a un sitio puede escindir el ácido nucleico diana unido por un ácido nucleico dirigido a ácido nucleico. El polipéptido dirigido a un sitio puede no escindir el ácido nucleico diana. En algunos casos, una endorribonucleasa escinde el ácido nucleico diana. La endorribonucleasa puede ser codificada por el vector. La endorribonucleasa puede ser codificada por el polinucleótido donante. La endorribonucleasa puede estar presente en la célula. La expresión de la endorribonucleasa y/o del polipéptido dirigido a un sitio puede ser inducida por un promotor condicional. Se puede incorporar un polinucleótido donante al ácido nucleico diana en el sitio donde se escindió.

Escisión

Los métodos desvelados en el presente documento pueden comprender además la escisión de todo, parte o nada del elemento indicador. Un primer ácido nucleico dirigido a ácido nucleico del elemento indicador puede dirigirse al extremo 5’ del elemento indicador. Un segundo ácido nucleico dirigido a ácido nucleico del elemento indicador puede dirigirse al extremo 3’ del elemento indicador. Un ácido nucleico dirigido a ácido nucleico puede dirigirse a los extremos 5' y 3' del elemento indicador. Un ácido nucleico dirigido a ácido nucleico puede dirigirse a dos secuencias en del elemento indicador y/o del polinucleótido donante. Las dos secuencias diana pueden ser al menos aproximadamente un 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % idénticas. Las dos secuencias diana pueden ser, como máximo, aproximadamente un 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % idénticas. Cuando se expresan los ácidos nucleicos dirigidos a ácido nucleico del elemento indicador, pueden formar un complejo con un polipéptido dirigido a un sitio y dirigirse a los extremos 5' y 3' del elemento indicador hibridándose a una región complementaria en los extremos 5’ y 3' del elemento indicador. La hibridación del complejo con el elemento indicador puede producir la escisión de todo, parte o nada del elemento indicador. El ácido nucleico escindido se puede volver a unir mediante, por ejemplo, unión de extremos no homólogos. El ácido nucleico unido de nuevo puede no introducir una eliminación o inserción. El ácido nucleico unido de nuevo puede introducir una eliminación o inserción. El ácido nucleico escindido se puede volver a unir mediante, por ejemplo, recombinación homóloga. La recombinación homóloga puede usarse para volver a unir un ácido nucleico escindido cuando los sitios del ácido nucleico diana son sustancialmente idénticos.

Cribado

Los métodos desvelados en el presente documento pueden comprender además la escisión de un elemento indicador de una célula seleccionada, formando así una segunda célula; y el cribado de la segunda célula. El cribado puede comprender la detección de la ausencia de todo o parte del elemento indicador. El cribado puede incluir la clasificación de células activadas por fluorescencia (FACS), en donde las células que expresan una proteína fluorescente codificada por el elemento indicador se separan de las células que no expresan una proteína fluorescente. Las células pueden ponerse en contacto con proteínas fluorescentes, sondas fluorescentes o anticuerpos conjugados con fluorocromo que se unen a proteínas codificadas por el elemento indicador o elemento genético, y posteriormente seleccionarse por FACS. Los fluorocromos pueden incluir, pero sin limitación, Cascade Blue, Pacific Blue, Pacific Orange, Lucifer yellow, NBD, R-Ficoeritrina (PE), conjugados de PE-Cy5, conjugados de PE-Cy7, Red 613, PerCP, TruRed, FluorX, Fluoresceína, BODIPY-FL, TRITC, Texas Red, Aloficocianina, conjugados de APC-Cy7 (PharRed), diversos colorantes Alexa Fluor, Cy2, Cy3, Cy3B, Cy3.5, Cy5, Cy5.5, Cy7, diversos DyLights, Y66H, Y66F, EBFP, EBFP2, Azurita, GFPuv, T-Zafiro, TagBFP, Ceruleano, mCFP, ECFP, CyPet, Y66W, dKeima-Red, mKeima-Red, TagCFP, AmCyan1, mTFP1, S65A, Midoriishi-Cyan, GFP, Turbo GFP, TagGFP, TagGFP2, AcGFP1, S65L, Esmeralda, S65T, S65C, EGFP, Azami-Green, ZaGreen1, Dronpa-Green, TagYFP, EYFP, Topaz, Venus, mCitirine, YPet, Turbo YFP, PhiYFP, PhiYFPm, ZaYellow1, mBanana, Kusabira-Orange, mOrange, mOrane2, mKO, TurboRFP, tdTomato, DsRed-Express2, TagRFP, monómero de DsRed, DsRed2, mStrawberry, Turbo FP602, AsRed2, mRFP1, J-Red, mCherry, HcRed1, mKate2, Katushka, mKate, TurboFP635, mPlum, mRaspberry, mNeptune, E2-Crimson.

Las células pueden ponerse en contacto con anticuerpos que se unen a marcadores de afinidad peptídicos codificados por el elemento indicador o elemento genético y, posteriormente, pueden seleccionarse mediante perlas inmunomagnéticas que reconocen los anticuerpos. El cribado puede comprender la tinción de células mediante la adición de X-gal cuando el elemento indicador o el elemento genético codifica b-galactosidasa. El cribado puede comprender una clasificación manual (por ejemplo, suspensiones celulares diluidas) y microscopía (por ejemplo, microscopía de fluorescencia). El cribado puede comprender un cribado de alto contenido.

Los elementos indicadores pueden codificar genes de resistencia a fármacos, lo que permite la selección de células que contienen el elemento indicador mediante la adición de fármacos, que destruyen las células que no expresan el elemento indicador. Dicho fármaco puede incluir, pero sin limitación, eritromicina, clindamicina, cloranfenicol, gentamicina, kanamicina, estreptomicina, tetraciclina, la combinación de quinupristina-dalfopristina, enrofloxacina, vancomicina, oxacilina, penicilina, sulfisoxazol sulfonamida, trimetoprima, sulfoximina metoinina, metotrexato, puromicina, blasticidina, histidinol, higromicina, zeocina, bleomicina y neomicina.

Bancos

La presente divulgación proporciona un banco de vectores de expresión que comprenden polinucleótidos donantes. En algunos aspectos, el banco puede comprender vectores de expresión que comprenden secuencias de polinucleótidos que codifican diferentes elementos genéticos de interés, pero los mismos elementos indicadores. En algunos aspectos, el banco puede comprender vectores de expresión que comprenden secuencias de polinucleótidos que codifican diferentes elementos genéticos de interés y diferentes elementos indicadores. Los elementos indicadores pueden diferir en sus secuencias de dirección de ácido nucleico (ARNcr y ARNcrtra). Los elementos indicadores pueden diferir en sus genes indicadores (por ejemplo, genes que codifican proteínas fluorescentes). La presente divulgación proporciona métodos para usar el banco para generar una pluralidad de células modificadas genéticamente. La presente divulgación proporciona métodos para usar el banco para un cribado genético de alto rendimiento. Estos bancos pueden permitir analizar grandes cantidades de genes individuales para inferir la función del gen. Los bancos pueden comprender de aproximadamente 10 miembros individuales a aproximadamente 10^{12} miembros individuales; por ejemplo, un banco puede comprender de aproximadamente 10 miembros individuales a aproximadamente 10^2 miembros individuales, de aproximadamente 10^2 miembros individuales a aproximadamente 10^3 miembros individuales, de aproximadamente 10^3 miembros individuales a aproximadamente 10^5 miembros individuales, de aproximadamente 10^5 miembros individuales a aproximadamente 10^7 miembros individuales, de aproximadamente 10^7 miembros individuales a aproximadamente 10^9 miembros individuales o de aproximadamente 10^9 miembros individuales a aproximadamente 10^{12} miembros individuales.

Modificación de células (transfección/infección)

Los métodos de la divulgación proporcionan la selección de células que comprenden el polinucleótido donante. En algunos aspectos, un método puede implicar poner en contacto un ácido nucleico diana o introducir en una célula (o una población de células) uno o más ácidos nucleicos que comprenden secuencias de nucleótidos que codifican un ácido nucleico dirigido a ácido nucleico, un polipéptido dirigido a un sitio y/o polinucleótido donante de la divulgación. Los métodos para introducir un ácido nucleico en una célula pueden incluir infección vírica o bacteriófaga, transfección, conjugación, fusión de protoplastos, lipofección, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina (PEI), transfección mediada con DEAE-dextrano, transfección mediada por liposomas, tecnología de pistola de partículas, precipitación con fosfato de calcio, microinyección directa, administración de ácido nucleico mediada por nanopartículas, y similares. En algunos aspectos, la puesta en contacto con un ácido nucleico diana o la introducción en una célula (o una población de células) de uno o más ácidos nucleicos puede no comprender una infección vírica. En algunos aspectos, la puesta en contacto con un ácido nucleico diana o la introducción en una célula (o una población de células) de uno o más ácidos nucleicos puede no comprender la infección por bacteriófagos. En algunos aspectos, la puesta en contacto con un ácido nucleico diana o la introducción en una célula (o una población de células) de uno o más ácidos nucleicos puede no comprender la transfección.

Ácidos nucleicos dirigidos a ácido nucleico diseñados

Dominios P diseñados

Se puede diseñar un ácido nucleico dirigido a ácido nucleico (por ejemplo, comprender modificaciones). Un ácido nucleico dirigido a ácido nucleico diseñado puede referirse a cualquiera de los ácidos nucleicos dirigidos a ácido nucleico diseñados como se describe en el presente documento. Por ejemplo, un ácido nucleico dirigido a ácido nucleico diseñado puede comprender una repetición CRISPR mínima, un ARNcrtra mínimo y un ARNcrtra 3'. Un dominio P de un ácido nucleico dirigido a ácido nucleico puede interactuar con la región de un polipéptido dirigido a un sitio. Un dominio P puede interactuar con una pluralidad de regiones de un polipéptido dirigido a un sitio. Un dominio P puede interactuar con una pluralidad de regiones de un polipéptido dirigido a un sitio, en donde al menos una de las regiones interactúa con un PAM en un motivo adyacente al protoespaciador. Los ejemplos de estas regiones pueden incluir los aminoácidos 1096-1225 y 1105-1138 de Cas9 en *S. pyogenes*.

Se puede introducir una modificación en el dominio P. Un dominio P puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o 30, o más, nucleótidos adyacentes. Un dominio P puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o 30, o más, nucleótidos adyacentes. Un dominio P puede iniciar un nucleótido 3’ del último nucleótido emparejado en el dúplex que comprende la repetición CRISPR mínima y la secuencia mínima de ARNcrtra. Un dominio P puede iniciar al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o 30, o más, nucleótidos a 3' del último nucleótido emparejado en el dúplex que comprende la repetición CRISPR mínima y la secuencia mínima de ARNcrtra. Un dominio P puede iniciar, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o 30, o más, nucleótidos a 3’ del último nucleótido emparejado en el dúplex que comprende la repetición CRISPR mínima y la secuencia mínima de ARNcrtra.

Un dominio P diseñado puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más mutaciones. Un dominio P diseñado puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, mutaciones. Las mutaciones pueden ser adyacentes entre sí (por ejemplo, secuenciales). Las mutaciones pueden estar separadas entre sí. Las mutaciones pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las mutaciones pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las mutaciones en un ácido nucleico dirigido a ácido nucleico pueden comprender inserciones, eliminaciones y sustituciones de nucleótidos en el ácido nucleico dirigido a ácido nucleico.

En algunos casos, un ácido nucleico dirigido a ácido nucleico diseñado comprende, como máximo, aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico dirigido a ácido nucleico de tipo silvestre. En algunos casos, un ácido nucleico dirigido a ácido nucleico diseñado comprende al menos aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico dirigido a ácido nucleico de tipo silvestre.

En algunos casos, una parte de ácido nucleico de CRISPR del ácido nucleico dirigido a ácido nucleico diseñado comprende, como máximo, aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico de CRISPR de tipo silvestre. En algunos casos, una parte de ácido nucleico de CRISPR del ácido nucleico dirigido a ácido nucleico diseñado comprende al menos aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico de CRISPR de tipo silvestre.

Una parte de ácido nucleico de ARNcrtra del ácido nucleico dirigido a ácido nucleico diseñado puede comprender, como máximo, aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico ARNcrtra de tipo silvestre. Una parte de ácido nucleico de ARNcrtra del ácido nucleico dirigido a ácido nucleico diseñado puede comprender al menos aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico ARNcrtra de tipo silvestre.

Las modificaciones en el dominio P pueden ser tales que el ácido nucleico dirigido a ácido nucleico diseñado se configure de nuevo para hibridarse con una nueva secuencia de PAM en un ácido nucleico diana. La modificación del dominio P puede ser complementaria a PAM en el ácido nucleico diana. La modificación del dominio P puede comprender el complemento inverso de PAM en el ácido nucleico diana. El nuevo PAM puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. El nuevo PAM puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las modificaciones en el dominio P pueden ocurrir en concierto con modificaciones en una región de unión al dominio P y unión a PAM de un polipéptido dirigido a un sitio. Estas modificaciones pueden ser compensatorias, en donde un dominio P modificado se modifica específicamente para unirse a un polipéptido dirigido a un sitio modificado, en donde la modificación permite que el polipéptido dirigido a un sitio se una al dominio P diseñado con mayor especificidad.

Un dominio P diseñado puede diseñarse para unirse a un nuevo PAM (por ejemplo, el dominio P diseñado puede hibridarse con un nuevo PAM). El nuevo PAM (por ejemplo, PAM diferente) puede ser bimodal (es decir, un PAM bimodal puede comprender dos regiones separadas del PAM). Las dos regiones separadas de un PAM bimodal pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las dos regiones separadas de un PAM bimodal pueden estar separadas, como máximo, por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos.

Un dominio P diseñado se puede diseñar para unirse a un nuevo PAM (por ejemplo, un PAM diferente), en donde el nuevo PAM es trimodal. Una PAM trimodal puede comprender tres regiones separadas de una secuencia de PAM (por ejemplo, tres regiones separadas que se pueden usar para dirigir un ácido nucleico dirigido a ácido nucleico a un ácido nucleico diana). Las tres regiones separadas de un PAM trimodal pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las tres regiones separadas de un PAM trimodal pueden estar separadas, como máximo, por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos.

Un ácido nucleico dirigido a ácido nucleico diseñado puede comprender al menos dos horquillas. Una primera horquilla puede comprender un dúplex entre la repetición CRISPR mínima y la secuencia mínima de ARNcrtra. La segunda horquilla puede estar cadena abajo de la primera horquilla. La segunda horquilla puede comenzar al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos cadena abajo del último nucleótido emparejado del primer dúplex. La segunda horquilla puede comenzar, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos cadena abajo del último nucleótido emparejado del primer dúplex. La segunda horquilla puede comprender un dominio P diseñado.

El dominio P diseñado de la segunda horquilla puede ubicarse en un lado de la horquilla del dúplex. El dominio P diseñado de la segunda horquilla puede ubicarse en ambos lados de la horquilla del dúplex. El dominio P diseñado de la segunda horquilla puede comprender al menos aproximadamente un 1, 2, 3, 4, 5, 10 o 20 % de los nucleótidos en la segunda horquilla. El dominio P diseñado de la segunda horquilla puede comprender, como máximo, aproximadamente un 1, 2, 3, 4, 5, 10 o 20 % de los nucleótidos en la segunda horquilla.

La segunda horquilla puede comprender un ARNcrtra (por ejemplo, el ARNcrtra medio o el ARNcrtra de 3' de un ácido nucleico dirigido a ácido nucleico). La segunda horquilla puede comprender al menos un 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 %, o más, de identidad con un ARNcrtra. La segunda horquilla puede comprender, como máximo, aproximadamente un 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 %, o más, de identidad con un ARNcrtra.

La segunda horquilla que comprende el dominio P diseñado puede configurarse para que elimine el dúplex (por ejemplo, se funda, se desenrolle). La segunda horquilla puede eliminar el dúplex cuando está en contacto con un ácido nucleico diana. La segunda horquilla puede eliminar el dúplex cuando está en contacto con un motivo adyacente al protoespaciador de un ácido nucleico diana.

En algunos casos, un dominio P modificado se puede configurar para hibridarse a una región en un ácido nucleico dirigido a ácido nucleico (por ejemplo, el mismo ácido nucleico dirigido a ácido nucleico que comprende el dominio P diseñado) y el dominio P diseñado se puede configurar para hibridarse con un ácido nucleico diana. En otras palabras, el dominio P diseñado puede comprender una secuencia conmutable, en la que, en algunos casos, el dominio P se hibrida con el ácido nucleico dirigido a ácido nucleico, formando así una horquilla, y, en algunos casos, el dominio P se hibrida a un PAM en un ácido nucleico diana.

Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para unirse a un PAM con una constante de disociación más baja que la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado (por ejemplo, ácido nucleico dirigido a ácido nucleico de tipo silvestre). Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado puede diseñarse para unirse a un PAM con una constante de disociación de al menos aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600 %, o más inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado.

Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado puede diseñarse para unirse a un PAM con una constante de disociación, como máximo, aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600%, o más inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado puede diseñarse para unirse a un PAM con una constante de disociación al menos aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, o más, inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado puede diseñarse para unirse a un PAM con una constante de disociación, como máximo, aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, o más, inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado.

Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para unirse a un PAM con una especificidad superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado (por ejemplo, ácido nucleico dirigido a ácido nucleico de tipo silvestre). Una mayor especificidad puede referirse a una reducción de la unión fuera de la diana (por ejemplo, la unión del ácido nucleico dirigido a ácido nucleico a una secuencia de PAM o de tipo PAM incorrecta). Por ejemplo, un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para reducir la unión no específica en al menos aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600 %, o más, en comparación con la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para reducir la unión no específica, como máximo, en aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600 %, o más, con respecto a la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para reducir la unión no específica, como máximo, aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, en comparación con la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende un dominio P modificado se puede diseñar para reducir la unión no específica en, como máximo, aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, o más, en comparación con la de un ácido nucleico dirigido a ácido nucleico que no comprende un dominio P modificado.

Protuberancias diseñadas

Se puede diseñar un ácido nucleico dirigido a ácido nucleico diseñado de manera que se pueda introducir una modificación en la región protuberante del ácido nucleico dirigido a ácido nucleico. Una protuberancia es un elemento típico del ácido nucleico que comprende nucleótidos no emparejados. Una protuberancia puede comprender nucleótidos no emparejados en cada cadena del dúplex que comprende la protuberancia. En otras palabras, una protuberancia puede comprender un nucleótido no emparejado en la cadena de repetición CRISPR mínima del dúplex y un nucleótido no emparejado en la cadena de secuencia mínima de ARNcrtra del dúplex.

Una protuberancia puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos no emparejados en una primera cadena de un dúplex en un ácido nucleico dirigido a ácido nucleico (es decir, la cadena de repetición CRISPR mínima del dúplex que comprende la repetición CRISPR mínima y la secuencia mínima de ARNcrtra). Una protuberancia puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos no emparejados en una primera cadena de un dúplex en un ácido nucleico dirigido a ácido nucleico (es

decir, cadena de repetición CRISPR mínima del dúplex que comprende la repetición mínima de CRISPR y la secuencia mínima de ARNcrtra). Una protuberancia puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos no emparejados en una segunda cadena de un dúplex en un ácido nucleico dirigido a ácido nucleico (es decir, la cadena de la secuencia mínima de ARNcrtra del dúplex que comprende la repetición mínima de CRISPR y la secuencia mínima de ARNcrtra). Una protuberancia puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos no emparejados en una segunda cadena de un dúplex en un ácido nucleico dirigido a ácido nucleico (es decir, la cadena de la secuencia mínima de ARNcrtra del dúplex que comprende la repetición mínima de CRISPR y la secuencia mínima de ARNcrtra). Una protuberancia puede comprender un nucleótido no emparejado en la secuencia mínima de ARN de CRISPR y 3 nucleótidos no emparejados en la cadena de secuencia mínima de ARNcrtra.

Los nucleótidos adyacentes a un nucleótido no emparejado pueden ser un nucleótido que forma una interacción de emparejamiento de bases oscilantes. Las interacciones de emparejamiento de bases oscilantes pueden incluir guanina-uracilo, hipoantina-uracilo, hipoxantina-adenina e hipoxantina-citosina. Al menos 1, 2, 3, 4 o 5 o más nucleótidos adyacentes a un nucleótido no emparejado pueden formar un emparejamiento oscilante. Como máximo, 1, 2, 3, 4 o 5, o más, nucleótidos adyacentes a un nucleótido no emparejado pueden formar un emparejamiento oscilante.

Una protuberancia diseñada puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, mutaciones. Una protuberancia diseñada puede comprender, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, mutaciones. Las mutaciones pueden ser adyacentes entre sí (por ejemplo, secuenciales). Las mutaciones pueden estar separadas entre sí. Las mutaciones pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las mutaciones pueden estar separadas por al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos. Las mutaciones en un ácido nucleico dirigido a ácido nucleico pueden comprender inserciones, eliminaciones y sustituciones de nucleótidos en el ácido nucleico dirigido a ácido nucleico.

Una protuberancia un ácido nucleico dirigido a ácido nucleico diseñado puede comprender, como máximo, aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 % o más de identidad y/o similitud de nucleótidos con un ácido nucleico dirigido a ácido nucleico de tipo silvestre. Una protuberancia un ácido nucleico dirigido a ácido nucleico diseñado puede comprender al menos aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 %, o más, de identidad y/o similitud de nucleótidos con un ácido nucleico dirigido a ácido nucleico de tipo silvestre.

Una cadena de la protuberancia puede estar mutada y la otra cadena no estar mutada. En otras palabras, en algunos casos, la secuencia de la protuberancia de la cadena mínima de ARN de CRISPR es la misma que la de un ácido nucleico dirigido a ácido nucleico de tipo silvestre, y la secuencia de la protuberancia de la secuencia mínima de ARNcr está mutada. En otras palabras, la secuencia de la protuberancia de la cadena mínima de ARN de CRISPR está mutada, y la secuencia de la protuberancia de la secuencia mínima de ARNcrtra es la misma que un ácido nucleico dirigido a ácido nucleico de tipo silvestre.

Las modificaciones en la protuberancia pueden ser tales que el ácido nucleico dirigido a ácido nucleico diseñado se configure de nuevo para unirse a un nuevo polipéptido dirigido a un sitio. El nuevo polipéptido dirigido a un sitio puede comprender al menos aproximadamente un 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 %, o más, de identidad de secuencia de aminoácidos con un polipéptido dirigido a un sitio de tipo silvestre. El nuevo polipéptido dirigido a un sitio puede comprender, como máximo, aproximadamente un 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 %, o más, de identidad de secuencia de aminoácidos con un polipéptido dirigido a un sitio de tipo silvestre. El nuevo polipéptido dirigido a un sitio puede ser un homólogo de Cas9. El nuevo polipéptido dirigido a un sitio puede ser un ortólogo de Cas9. El nuevo polipéptido dirigido a un sitio puede ser una quimera de dos polipéptidos diferentes dirigidos a un sitio. El nuevo polipéptido dirigido a un sitio puede comprender una mutación como se desvela en el presente documento.

Un ácido nucleico dirigido a ácido nucleico diseñado que comprende una protuberancia modificada puede diseñarse para unirse a un polipéptido dirigido a un sitio con una constante de disociación inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende una protuberancia modificada (por ejemplo, ácido nucleico dirigido a ácido nucleico de tipo silvestre). Un ácido nucleico dirigido a ácido nucleico diseñado que comprende una protuberancia modificada puede diseñarse para unirse a un polipéptido dirigido a un sitio con una constante de disociación de al menos aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600 %, o más, inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende la protuberancia modificada. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende una protuberancia modificada puede diseñarse para unirse a un polipéptido dirigido a un sitio con una constante de disociación, como máximo, aproximadamente un 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 o 600 %, o más inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende una protuberancia modificada. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende una protuberancia modificada puede diseñarse para unirse a un polipéptido dirigido a un sitio con una constante de disociación al menos aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, o más, inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende una protuberancia modificada. Un ácido nucleico dirigido a ácido nucleico diseñado que comprende una protuberancia modificada puede diseñarse para unirse a un polipéptido dirigido

a un sitio con una constante de disociación, como máximo, aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces o 50 veces, o más, inferior o superior a la de un ácido nucleico dirigido a ácido nucleico que no comprende una protuberancia modificada.

Métodos

La divulgación proporciona métodos para diseñar ácidos nucleicos dirigidos a ácido nucleico. Los métodos pueden comprender la modificación de un ácido nucleico dirigido a ácido nucleico. La modificación puede comprender insertar, eliminar, sustituir y mutar los nucleótidos en el ácido nucleico dirigido a ácido nucleico. La modificación puede comprender modificar el ácido nucleico dirigido a ácido nucleico de manera que el ácido nucleico dirigido a ácido nucleico pueda unirse a nuevos motivos adyacentes al protoespaciador y/o nuevos polipéptidos dirigidos a un sitio en comparación con un ácido nucleico dirigido a ácido nucleico de tipo silvestre. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

Se puede usar un ácido nucleico dirigido a ácido nucleico diseñado para escindir un ácido nucleico diana. Un ácido nucleico dirigido a ácido nucleico diseñado puede introducirse en células con un polipéptido dirigido a un sitio, formando así un complejo. El complejo puede hibridarse con un ácido nucleico diana, en donde el ácido nucleico diana comprende un motivo adyacente al protoespaciador. El polipéptido dirigido a un sitio del complejo puede escindir el ácido nucleico diana.

Las partes complementarias de las secuencias de ácido nucleico del ácido nucleico pre-CRISPR y las secuencias de ácido nucleico crtra de *Streptococcus pyogenes* SF370 se muestran en la **Figura 20**.

La **Figura 21** representa una estructura ilustrativa del dúplex (por ejemplo, horquilla) que comprende la repetición CRISPR mínima y la secuencia mínima de ARNcrtra y una parte de la secuencia de ARNcrtra de 3'. El dúplex comprende una región de protuberancia.

La Tabla 4 contiene las secuencias de plantillas de ADN usadas para sintetizar los ácidos nucleicos dirigidos a un ácido nucleico guía sencillos de la divulgación.

Tabla 4. Plantillas de ADN de los ácidos nucleicos guía de la divulgación

Grupo 1 de variantes de dúplex

	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAG
	GCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT
TEMP3-FL	T
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGAAAA
	AGAGCTAGAAATAGCAAGTTTTTTTAAGGCTAGTCCGTTATCAACTTGA
SGR-v2	AAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGATAT
	AGAGCTAGAAATAGCAAGTTATATTAAGGCTAGTCCGTTATCAACTTG
SGR-v3	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGGATGAAAATCCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v4	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGAAAA
	TGAGGATGAAAATCCAAGTATTTTTTAAGGCTAGTCCGTTATCAACTTGA
SGR-v5	AAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGATTA
	TGAGGATGAAAATCCAAGTATAATTAAGGCTAGTCCGTTATCAACTTG
SGR-v6	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(continuación)

Grupo 1 de variantes de dúplex

	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTAAT
	TGAGGATGAAAATCCAAGTAATTATAAGGCTAGTCCGTTATCAACTTG
SGR-v7	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGAAAA
	TCAAGTGATGAAAATCGAGATTTTTAAGGCTAGTCCGTTATCAACTTGA
SGR-v8	AAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGAAAA
	TGAAGGATGAAAATCCAGTATTTTTAAGGCTAGTCCGTTATCAACTTGA
SGR-v9	AAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGATTT
	AGAGCTAGAAATAGCAAGTTAAATTAAGGCTAGTCCGTTATCAACTTG
SGR-v10	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTCTC
	AGAGCTAGAAATAGCAAGTTGAGATAAGGCTAGTCCGTTATCAACTTG
SGR-v11	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTCCC
	AGAGCTAGAAATAGCAAGTTGGGATAAGGCTAGTCCGTTATCAACTTG
SGR-v12	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGACTCAGAAATCAGAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v13	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(continuación)

Grupo 2 de variantes de dúplex	
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCTCTAAAATAAGGCTAGTCCGTTATCAACTTGAA
SGR-v14	AAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT
	AGAGGAAACTCTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
SGR-v15	GCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT
	AGAAATAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG
SGR-v16	AGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATATTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v17	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATATTTT
	AGAGCTAGAAATAGCAAGTTAAAACAAGGCTAGTCCGTTATCAACTTG
SGR-v18	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGACGA
	TAGAACGGAAACGTTGGACATCGTTAAGGCTAGTCCGTTATCAACTTG
SGR-v19	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGACGA
	TGAGACGGAAACGTCAAGTATCGTTAAGGCTAGTCCGTTATCAACTTG
SGR-v20	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(continuación)

Grupo 2 de variantes de dúplex

	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AAGACTAGAAATAGTGGACTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v21	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATCGTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v22	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTGGT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v23	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTG
	CGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
SGR-v24	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGTGAGAAATAGCAAGTTCACATAAGGCTAGTCCGTTATCAACTTG
SGR-v25	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTACACTAAGGCTAGTCCGTTATCAACTTG
SGR-v26	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAACAGAAGGCTAGTCCGTTATCAACTTG
SGR-v27	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(continuación)

Grupo 2 de variantes de dúplex

AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT

AGAGCTAGAAATAGCAAGTTAAAATAACTGGCTAGTCCGTTATCAACT

SGR-v28 TGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT

AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT

AGAGCTAGAAATAGCAAGTTAAAATGCTAGTCCGTTATCAACTTGAAA

SGR-v29 AAGTGGCACCGAGTCGGTGCTTTTTTTT

Grupo de variantes de crtra

AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT

AGAGCTAGAAATAGCAAGTTAAAATGGAACTAGTCCGTTATCAACTTG

SGR-v30 AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT

AGAGCTAGAAATAGCAAGTTAAAATTTCGCTAGTCCGTTATCAACTTG

SGR-v31 AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

AGTAATAATACGACTCACTATAGGGGGGCCACTAGGGACAGGATGTTTT

AGAGCTAGAAATAGCAAGTTAAAATAAGCGAAGTCCGTTATCAACTTG

SGR-v32 AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(continuación)

Grupo de variantes de crtra	
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTTCACCGTTATCAACTTG
SGR-v33	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTGGCTTATCAACTTG
SGR-v34	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGAATTCAACTTG
SGR-v35	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAAGTACTTG
SGR-v36	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCATGATG
SGR-v37	AAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGAATGATACATCACAAAAAA
	AAGGCTTTATGCCGTAACTACTACTTATTTTCAAAATAAGTAGTTTTTTT
SGR-v38-MMO	T

(continuación)

Grupo de variantes de crtra	
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTTCATGCCGAAATCAACA
SGR-v39-ST2	CCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAAGAGTTTGCGGGACTCTGCGG
SGR-v40-CJ	GGTTACAATCCCCTAAAACCGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCCGTCTGAAAAGATGTGCC
SGR-v41-NM	GCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATTTTTT
Grupo de marcadores de Csy4	
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTA
SGR-v42	TACTGCCGTATAGGCAGAGATTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTA
	TACTGCCGTATAGGCAGAGAAATGGACTCGGAATACTGCCGTATAGGC
SGR-v43	AGAGATTTTTT

(continuación)

Grupo de marcadores de Csy4	
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
SGR-v44	AGAGCTATCACTGCCGTATAGGCAGTGATAGCAAGTTAAAATAAGGCT
	AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	AGTAATAATACGACTCACTATAGGGGGCCACTAGGGACAGGATGTTTT
	AGAGCTATCACTGCCGTATAGGCAGTGATAGCAAGTTAAAATAAGGCT
	AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTAATGGAC
SGR-v45	TCGATACTGCCGTATAGGCAGAGATTTTTT

La Tabla 4A muestra las secuencias de ARN de las plantillas de ADN de la Tabla 4.

Tabla 4A. Secuencias de ARN de los ácidos nucleicos dirigidos a ácido nucleico guía sencillos de la divulgación

P/N	Secuencia de ácido nucleico dirigido a ácido nucleico guía sencillo
Grupo 1 de variantes de dúplex	
TEMP3-FL	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAUGCUGUUUUGGAAACA
	AAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAA
	AAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v2	GGGGCCACUAGGGACAGGAUGAAAAAAGAGCUAGAAAUAGCAAGUUUUU
	UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
	UUUUUUU
SGR-v3	GGGGCCACUAGGGACAGGAUGAUAUAGAGCUAGAAAUAGCAAGUUAUA
	UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
	UUUUUUU
SGR-v4	GGGGCCACUAGGGACAGGAUGUUUUAGAGGAUGAAAAAUCCAAGUUAAA
	AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
	UUUUUUU
SGR-v5	GGGGCCACUAGGGACAGGAUGAAAAAUGAGGAUGAAAAAUCCAAGUAUUU
	UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
	UUUUUUU
SGR-v6	GGGGCCACUAGGGACAGGAUGAUUAUGAGGAUGAAAAAUCCAAGUAUAA
	UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
	UUUUUUU

(continuación)

P/N	Secuencia de ácido nucleico dirigido a ácido nucleico guía sencillo
Grupo 1 de variantes de dúplex	
SGR-v7	GGGGCCACUAGGGACAGGAUGUAAUUGAGGAUGAAAAUCCAAGUAAUU AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v8	GGGGCCACUAGGGACAGGAUGAAAAUCAAGUGAUGAAAAUCGAGAUUU UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v9	GGGGCCACUAGGGACAGGAUGAAAAUGAAGGAUGAAAAUCCAGUAUUU UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v10	GGGGCCACUAGGGACAGGAUGAUUUAGAGCUAGAAAUAGCAAGUUAAA UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v11	GGGGCCACUAGGGACAGGAUGUCUCAGAGCUAGAAAUAGCAAGUUGAG AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v12	GGGGCCACUAGGGACAGGAUGUCCCAGAGCUAGAAAUAGCAAGUUGGG AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU

(continuación)

P/N	Secuencia de ácido nucleico dirigido a ácido nucleico guía sencillo
Grupo 1 de variantes de dúplex	
SGR-v13	GGGGCCACUAGGGACAGGAUGUUUUAGACUCAGAAAUCAGAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
Grupo 2 de variantes de dúplex	
SGR-v14	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCUCUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
SGR-v15	GGGGCCACUAGGGACAGGAUGUUUUAGAGGAAACUCUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v16	GGGGCCACUAGGGACAGGAUGUUUUAGAAAUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v17	GGGGCCACUAGGGACAGGAUAUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v18	GGGGCCACUAGGGACAGGAUAUUUUAGAGCUAGAAAUAGCAAGUUAAAACAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

(continuación)

Grupo 2 de variantes de dúplex

SGR-v19	GGGGCCACUAGGGACAGGAUGACGAUAGAACGGAAACGUUGGACAUCG UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v20	GGGGCCACUAGGGACAGGAUGACGAUGAGACGGAAACGUCAAGUAUCG UUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v21	GGGGCCACUAGGGACAGGAUGUUUUAAGACUAGAAAUAGUGGACUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v22	GGGGCCACUAGGGACAGGAUCGUUUAGAGCUAGAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v23	GGGGCCACUAGGGACAGGAUGUGGUAGAGCUAGAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU
SGR-v24	GGGGCCACUAGGGACAGGAUGUUUGCGAGCUAGAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUUUU

(continuación)

Grupo 2 de variantes de dúplex

SGR-v25	GGGGCCACUAGGGACAGGAUGUUUUAGAGUGAGAAAUAGCAAGUUCACAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v26	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUACACUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v27	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAACAGAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v28	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAACUGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v29	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

(continuación)

Grupo de variantes de crtra	
SGR-v30	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUGGAACUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v31	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUUUCGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v32	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGCGAAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v33	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUUCACCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v34	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUGGCUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v35	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGAAUUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

(continuación)

Grupo de variantes de crtra	
SGR-v36	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAAGUACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v37	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAUGAUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v38-MMO	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGAAUGAUACAUCACAAAAAAAAGGCUUUAUGCCGUAACUACUACUUAUUUUCAAAAUAAGUAGUUUUUUUU
SGR-v39-ST2	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUUCGUUAUUUUUUU
SGR-v40-CJ	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUUUUU
SGR-v41-NM	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGGGCAUUUUUU

(continuación)

Grupo de marcadores de Csy4	
SGR-v42	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUAUACUGCCGUAUAGGCAGAGAUUUUUU
SGR-v43	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUAUACUGCCGUAUAGGCAGAGAAAUGGACUCGGAAUACUGCCGUAUAGGCAGAGAUUUUUU
SGR-v44	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAUCACUGCCGUAUAGGCAGUGAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
SGR-v45	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAUCACUGCCGUAUAGGCAGUGAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUAAUGGACUCGAUACUGCCGUAUAGGCAGAGAUUUUUU

La Tabla 5 indica la actividad y el propósito de variantes adicionales de ácido nucleico dirigido a ácido nucleico. + se refiere a activo, - se refiere a inactivo. Los datos experimentales de estas variantes se muestran en la **Figura 37**.

Tabla 5. Variantes de ácido nucleico dirigido a ácido nucleico y su actividad

nombre	**n.º de dCB**	**Actividad**	**Propósito**	**Espaciador**	**Estructura principal gs**
SGR-v80		-	Longitud del enlazador	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAACTCG GCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGT CGGTGCT
SGR-v81		-	Longitud del enlazador	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAACTCT GGCTAGTCCGTTATCAACT TGAAAAAGTGGCACCGAG TCGGTGCT
SGR-v82		-	Longitud del enlazador	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAACTCT CTGGCTAGTCCGTTATCAA CTTGAAAAAGTGGCACCG AGTCGGTGCT
SGR-v83		++	Control	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGG TGCT
SGR-v84		++	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGGAAACAAGT TAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCT
SGR-v85		++	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGACAAGTTAA AATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCA CCGAGTCGGTGCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v86		++	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTAGGAGAAACTTTAA AATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCA CCGAGTCGGTGCT
SGR-v87		-	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTATCGAAATCTAAAA TAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACC GAGTCGGTGCT
SGR-v88		-	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTACTTCGGTAAAATA AGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGA GTCGGTGCT
SGR-v89		++	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTAGATACTTAAAATA AGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGA GTCGGTGCT
SGR-v90		-	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTTATGAAACTAAAATA AGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGA GTCGGTGCT
SGR-v91		-	ARNgs mínimo	GGGGC CACTA GGGAC AGGAT	GTTTCTTCGGAAATAAGGC TAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCG GTGCT
SGR-v92		++	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGGCTAGAAATAG CAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGT GCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v93		++	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGCTAGAAATAGC AAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG CT
SGR-v94		++	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTACTAGAAATAGCA AGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG CT
SGR-v95		-	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGT GCT
SGR-v96		-	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG CT
SGR-v97		++	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA AGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG CT
SGR-v98		-	Cambio de ángulo de la protuberancia	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGC T

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v99		-	Control	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGG TGCT
SGR-v124	dCB154	-	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAACCCT AGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGG TGCT
SGR-v125	dCB155	+/-	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAACCCT AGTGGGTTATCAACTTGA AAAAGTGGCACCGAGTCG GTGCT
SGR-v126	dCB156	+	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTGGGTTATCAACTTGA AAAAGTGGCACCGAGTCG GTGCT
SGR-v46	dCB157	++	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTAAGTTATCAACTTGA AAAAGTGGCACCGAGTCG GTGCT
SGR-v47	dCB158	++	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTTTGTTATCAACTTGAA AAAGTGGCACCGAGTCGG TGCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v48	dCB159	++	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGACT AGTTCGTTATCAACTTGAA AAAGTGGCACCGAGTCGG TGCT
SGR-v49	dCB160	-	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGGAAACTAGCCTTCTCA A
SGR-v50	dCB161	-	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGGAAACTAGCCCTCAAT A
SGR-v51	dCB162	+/-	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTAGAAAATGCC
SGR-v52	dCB163	+	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTAGAAATACGG
SGR-v53	dCB164	+	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTAGAAATACTTAT
SGR-v54	dCB165	+	variante de nexo/horquilla	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTTATTATTAGGGG GTTA

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v55	dCB166	-	Truncamiento del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGG
SGR-v56	dCB167	-	Truncamiento del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AG
SGR-v57	dCB168	+/-	truncamiento del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCG
SGR-v58	dCB169	+/-	truncamiento del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT AGTCCGTTAT
SGR-v100		+/-	longitud del tallo del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGGC TAGTCCCGTTATCAACTTG AAAAAGTGGCACCGAGTC GGTGCT
SGR-v101		++	longitud del tallo del nexo	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGGG CTAGTCCCCGTTATCAACT TGAAAAAGTGGCACCGAG TCGGTGCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v102		+	longitud del tallo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGGGGCTAGTCCCCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v103		++	longitud del tallo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAAACTAGTTTGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v104		++	longitud del tallo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGACTAGTTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v105		++	Control	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v106		+/-	longitud del tallo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCTGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v107		++	longitud del tallo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAAGCTAGTCTGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v108		-	tamaño del lazo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGTCTAGTCCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v109		+	tamaño del lazo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGTACTAGTCGCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v110		-	tamaño del lazo del nexo	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGTATCTAGTGCGCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v111		++	Control	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SGR-v112		-	Híbrido de crtra - LRH	GGGGCCACTAGGGACAGGAT	GTTTTAGAGCTAGAAATAGCAAGTTAAAATCAAACAAAGCTTCAGCTGAGTTTCAATTTCTGGCCCATGTTGGGCACATACATATGCCACCGAG

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v113		++	Híbrido de crtra - SM159	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCA GTGATTTTTAATCCAGTCC GTACACAACTTGAAAAAG TGCGCACCGATTCGGTGCT TTTTT
SGR-v114		++	Híbrido de crtra - ST3	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCT TAGACCGTACTCAACTTGA AAAGGTGGCACCGATTCG GTGTTTTTTTT
SGR-v115		++	Híbrido de crtra - LBU	GGGGC CACTA GGGAC AGGAT	GTTTTAGAGCTAGAAATA GCAAGTTAAAATCAAAGC GCTTTGCGCGGAGTTTCAA CTTTT
SGR-v120		-	Variante principal	GGGGC CACTA GGGAC AGGAT	GAGAATCTCCTAGAAATA GCTCTTATTCTTAAGGGAT CACCGAATACAACTTGAA AAAGTGGCACCGAGTCGG TGCT
SGR-v121		++	Variante principal	GGGGC CACTA GGGAC AGGAT	GACGATGAGACGGAAACG TCAAGTATCGTTAAGGGA TCACCGAATACAACTTGA AAAAGTGGCACCGAGTCG GTGCT
SGR-v122		-	Variante principal	GGGGC CACTA GGGAC AGGAT	AACGATGAGACGGAAACG TCAAGTATCGTCAAGGGA TCACCCAATACAACTTGA AAAAGTGGCACCGAGTCG GTGCT

(continuación)

nombre	n.º de dCB	Actividad	Propósito	Espaciador	Estructura principal gs
SGR-v123		-	Variante principal	GGGGC CACTA GGGAC AGGAT	AACGGTGAGGTGGAAACA CCAAGTACCGTCAAGGTA GCACCCGACAAGTC

Métodos para la generación de estirpes celulares marcadas usando un ácido nucleico dirigido a ácido nucleico

Los métodos de la divulgación proporcionan el marcaje de una célula con un polinucleótido donante, en donde el polinucleótido donante puede dividirse y/o diferenciarse, y el polinucleótido donante puede transmitirse a cada célula hija durante la división celular. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

Se puede generar una célula marcada poniendo en contacto la célula con un polinucleótido donante y un complejo que comprenda un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico. El polinucleótido donante puede insertarse en el ácido nucleico diana escindido, generando así una célula marcada. La célula marcada se puede propagar, tal como en una estirpe celular, o para producir una población propagada de células.

Se puede introducir un polinucleótido donante en el sitio de corte mediante el uso de un casete donante para la recombinación homóloga que comprende extremos homólogos a secuencias a ambos lados de la rotura bicatenaria. El polinucleótido donante puede comprender una secuencia adicional entre los dos extremos. La secuencia adicional puede ser una secuencia de ácido nucleico. La secuencia adicional puede codificar un gen. La secuencia adicional puede codificar un elemento de ácido nucleico no codificante.

El polinucleótido donante (por ejemplo, la secuencia adicional de un polinucleótido donante entre dos extremos homólogos) puede comprender un marcador. Un marcador puede comprender un marcador de visualización (por ejemplo, un marcador fluorescente tal como GFP). Un marcador puede comprender una secuencia de polinucleótido aleatoria (por ejemplo, tal como una secuencia de hexámero aleatoria). Un marcador puede ser un código de barras.

La NHEJ puede introducir una firma de secuencia única en cada sitio de corte. El mecanismo de reparación puede producir la introducción de inserciones (por ejemplo, inserción de un polinucleótido donante), eliminaciones o mutaciones en un sitio de corte. Una célula que se somete a NHEJ para reparar una rotura bicatenaria puede comprender una secuencia única una vez que haya tenido lugar la reparación (por ejemplo, se puede insertar una secuencia única en la rotura bicatenaria). Si se corta más de un sitio dentro de una célula, la reparación puede introducir el polinucleótido donante en cada sitio, agregando así diversidad de secuencia a esa célula. El sitio reparado puede proporcionar una secuencia de código de barras única a la célula, que puede conservarse durante la división celular y transmitirse a toda la progenie de la célula modificada. Se puede insertar un polinucleótido donante en al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más sitios (por ejemplo, ácidos nucleicos diana escindidos). Se puede insertar un polinucleótido donante, como máximo, en aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más sitios (por ejemplo, ácidos nucleicos diana escindidos).

La recombinación homóloga (HR) puede usarse para introducir secuencias de códigos de barras en una célula y/o una población de células (por ejemplo, una célula humana, una célula de mamífero, una levadura, un hongo, un protozoo, una arquea). Se puede preparar un banco de plásmidos donantes (por ejemplo, que comprenden el polinucleótido donante) con secuencias aleatorizadas en el casete donante. El banco puede estar hecho de oligonucleótidos, una pieza de ADN bicatenario, un plásmido y/o un minicírculo.

Las secuencias de polinucleótidos donantes pueden introducirse en los genomas de células individuales con el fin de rastrear el linaje celular. Los sitios se pueden elegir para su modificación en regiones silenciosas o de "puerto seguro" del genoma, distantes de los genes y elementos reguladores, para minimizar los efectos potencialmente nocivos sobre la función celular. Los sitios dentro de elementos genéticos funcionales también se pueden usar para rastrear el destino celular.

Por ejemplo, los polinucleótidos donantes pueden introducirse en células madre y/o poblaciones de células madre. Los métodos de la divulgación se pueden usar para rastrear el desarrollo del linaje celular en modelos animales. Por ejemplo, se puede rastrear el desarrollo del destino celular y/o la diferenciación en la hematopoyesis usando los métodos de la divulgación. Los métodos de la divulgación pueden usarse para terapias basadas en ingeniería de células terapéuticas. Por ejemplo, una célula puede marcarse con un polinucleótido donante que codifique una

proteína terapéutica. La célula se puede propagar. La célula propagada se puede introducir en un sujeto. Como otro ejemplo, una célula diferenciada puede retirarse de un sujeto. La célula diferenciada puede marcarse con dos marcadores: uno expresado cuando la célula está diferenciada, otro expresado cuando la célula está desdiferenciada. La identificación de los marcadores puede ser útil para determinar los eventos de diferenciación. En otro ejemplo, se puede obtener una célula diferenciada de un sujeto. La célula diferenciada se puede desdiferenciar en una célula pluripotente. La célula pluripotente puede marcarse con un polinucleótido donante que codifique una proteína terapéutica. La célula se puede volver a diferenciar en un nuevo tipo celular mientras se expresa la proteína terapéutica, creando así una célula terapéutica específica del paciente. Las células marcadas pueden dividirse y diferenciarse, y las modificaciones realizadas en su genoma pueden transmitirse a cada célula hija durante la división celular.

En algunos casos, se pueden marcar dos células con dos marcadores de polinucleótido donante diferentes. Las dos células se pueden combinar. La mezcla combinada se puede analizar simultáneamente. Los polinucleótidos donantes pueden permitir el análisis multiplex de las dos células, porque el polinucleótido donante puede usarse para distinguir las dos células.

Se puede elegir una población celular para introducir roturas bicatenarias o generar firmas celulares. Las células pueden ser purificadas o seleccionadas. Por ejemplo, una población de células madre hematopoyéticas (CD45 positivas) puede seleccionarse mediante FACS o purificación de perlas magnéticas. La médula ósea se puede tratar ex vivo con la nucleasa. Las células pueden ser atacadas *in vivo* mediante el uso de virus con un tropismo particular. Las células pueden seleccionarse mediante el uso de virus diseñados para dirigirse a células que portan un receptor particular.

Las células marcadas pueden analizarse mediante secuenciación de alto rendimiento a nivel de población o a nivel de célula individual. A nivel de población, se puede lisar una colección de células. El ADN genómico se puede extraer. Se pueden diseñar cebadores de PCR para amplificar la región genómica que ha sido modificada por la nucleasa. Las secuencias pueden enriquecerse por hibridación. Se puede preparar un banco de secuencias a partir del ADN genómico y enriquecerla. Se puede enriquecer la región de interés y se puede preparar un banco de secuencias. Se puede preparar un banco de secuencias simultáneamente durante el enriquecimiento usando cebadores que comprendan marcadores de secuencia apropiados para su uso con tecnologías de secuenciación de ácido nucleico. Si la rotura bicatenaria se realiza dentro de una región que se puede transcribir, se puede usar ARN para preparar bancos de secuencias.

Una vez que se han obtenido los datos de la secuencia de ácido nucleico, se pueden analizar las secuencias para determinar la estructura clónica. Esto puede llevarse a cabo reuniendo secuencias comunes y contando esas secuencias.

Las células se pueden subseleccionar mediante esquemas de clasificación basados en marcadores de superficie celular usando citometría de flujo o métodos de purificación por afinidad. Los marcadores de superficie celular se pueden usar para definir estados celulares y, si se comparan los estados celulares con la estructura clónica, se puede determinar el destino de las poblaciones celulares modificadas.

A nivel de una sola célula, las células pueden aislarse. Los productos de PCR se pueden generar a partir de cada célula individual. Esto se puede lograr en matrices de micropocillos, dispositivos microfluídicos y/o emulsiones. Cuando se realiza más de una modificación genómica por célula, los productos de PCR se pueden acoplar entre sí, ya sea física o químicamente, para garantizar su relación con la célula madre.

Métodos para cuantificar eventos de edición del genoma

Para las nucleasas dependientes del ARN, tales como Cas9, se pueden unir la funcionalidad de reconocimiento de ácido nucleico y las actividades nucleasa. En algunos casos, la funcionalidad de reconocimiento de ácido nucleico y las actividades nucleasa pueden no estar enlazadas. Los sitios de nucleasa pueden ubicarse dentro de la secuencia específica reconocida por la nucleasa.

La unión de extremos no homólogos puede ser un proceso de reparación imperfecto que puede producir en la inserción de múltiples bases en el sitio de la rotura bicatenaria. La NHEJ puede producir la introducción de inserciones, eliminaciones y/o mutaciones en un sitio de corte. La NHEJ puede interrumpir significativamente la secuencia original. Se puede usar la interrupción de la secuencia nativa como consecuencia de los mecanismos de reparación para evaluar la eficiencia de los enfoques de edición del genoma.

La recombinación homóloga puede permitir una reparación más completa de la rotura del ácido nucleico diana mediante el intercambio de secuencias de nucleótidos entre moléculas de ácido nucleico similares o idénticas. Se puede introducir una secuencia adicional en el ácido nucleico diana en el sitio de corte mediante el uso de un casete donante (por ejemplo, polinucleótido donante) que comprenda extremos homólogos a las secuencias a ambos lados de la rotura bicatenaria y una secuencia adicional entre los dos extremos.

La presente divulgación describe un enfoque para evaluar la actividad de rotura bicatenaria y las inserciones/eliminaciones mediadas por la NHEJ introducidas por nucleasas dependientes de ácido nucleico, tales como Cas9. El método aprovecha el hecho de que los sitios de un ácido nucleico diana reconocido por Cas9 durante el reconocimiento inicial de la nucleasa y la actividad de escisión de ácido nucleico pueden destruirse durante el proceso de NHEJ, ya sea mediante la introducción de inserciones o eliminaciones.

En algunos casos, el método proporciona el diseño de un ácido nucleico dirigido a ácido nucleico para dirigir un sitio de interés en un ácido nucleico diana (por ejemplo, genoma). Se puede diseñar una plantilla de ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico con una secuencia promotora añadida en el extremo 5' del ácido nucleico dirigido a ácido nucleico para permitir la síntesis *in vitro* del ácido nucleico dirigido a ácido nucleico.

Los cebadores pueden diseñarse en posiciones que flanquean el sitio de escisión. El sitio de escisión (y/o regiones de ácido nucleico alrededor del sitio de escisión) puede amplificarse (por ejemplo, a partir de ácido nucleico genómico), generando así un producto (por ejemplo, producto de PCR amplificado). El producto (por ejemplo, producto de PCR amplificado) puede tener al menos aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200 o más bases de longitud. El producto (por ejemplo, producto de PCR amplificado) puede tener, como máximo, aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200 o más bases de longitud. El producto (por ejemplo, producto de PCR amplificado) puede tener aproximadamente 200-600 pares de bases de longitud.

Los productos pueden ser purificados. Los productos pueden ser incubados con una nucleasa dependiente del ARN (por ejemplo, Cas9) y ácido nucleico dirigido a ácido nucleico. Esas moléculas que han sido amplificadas a partir de ácido nucleico genómico que no han sido modificadas por NHEJ pueden comprender la secuencia correcta que puede ser reconocida y escindida por Cas9. Las moléculas que se han amplificado a partir del ácido nucleico genómico que ha sido modificado por NHEJ pueden no comprender sitios que puedan ser reconocidos y/o cortados por Cas9.

Los productos digeridos pueden analizarse luego mediante métodos tales como electroforesis en gel, electroforesis capilar, secuenciación de alto rendimiento y/o PCR cuantitativa (por ejemplo, qPCR). En el caso de la electroforesis en gel, se puede obtener una imagen de un gel. Una vez que se ha tomado una imagen de un gel, se puede estimar el porcentaje de células modificadas por NHEJ midiendo la intensidad de las bandas correspondientes a los productos digeridos y comparando la intensidad de las bandas correspondientes a los productos no digeridos.

Métodos para administrar el polinucleótido donante a una rotura bicatenaria para su inserción en la rotura bicatenaria

La presente divulgación describe métodos para acercar un polinucleótido donante a una rotura de ácido nucleico diana dirigido a un sitio para mejorar la inserción (por ejemplo, recombinación homóloga) del polinucleótido donante en el sitio de la rotura bicatenaria. El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

En algunos casos, los métodos de la divulgación prevén llevar un polinucleótido donante cerca del sitio de una rotura bicatenaria en un ácido nucleico diana, uniéndolo a la nucleasa que genera la rotura bicatenaria (por ejemplo, Cas9).

Un complejo que comprende un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y un polinucleótido donante pueden administrarse a un ácido nucleico diana. La **Figura 30** ilustra métodos ilustrativos para acercar un polinucleótido donante a un sitio de una rotura bicatenaria en un ácido nucleico diana. Por ejemplo, un ácido nucleico dirigido a ácido nucleico puede comprender una secuencia de extensión de hibridación de 3', que puede ser parte de una secuencia de extensión de ARNcrtra (mostrada en la línea punteada clara unida al ácido nucleico dirigido a ácido nucleico). Una secuencia de extensión de hibridación de 3' puede ser una secuencia no nativa. La **Figura 30A** ilustra que la secuencia de extensión de ARNcrtra en el extremo 3' del ácido nucleico dirigido a ácido nucleico puede incluir una secuencia que puede hibridarse con un extremo del polinucleótido donante (por ejemplo, el extremo 3') (el polinucleótido donante se muestra en línea punteada gruesa en negrita). La secuencia de extensión de hibridación de 3' puede tener al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos de longitud. La secuencia de extensión de hibridación de 3' puede tener, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos de longitud. La secuencia de hibridación de 3' puede hibridarse con al menos aproximadamente 1, 2, 3, 4, 5, 5, 6, 7, 8, 9 o 10, o más, nucleótidos del polinucleótido donante. La secuencia de hibridación de 3' puede hibridarse, como máximo, con 1, 2, 3, 4, 5, 5, 6, 7, 8, 9 o 10, o más, nucleótidos del polinucleótido donante. La secuencia de hibridación de 3' puede hibridarse con el polinucleótido donante con al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, emparejamientos erróneos. La secuencia de hibridación de 3' puede hibridarse con el polinucleótido donante con, como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más emparejamientos erróneos.

La extensión de hibridación de 3' puede hibridarse con el extremo 3' del polinucleótido donante. La extensión de hibridación de 3' puede hibridarse con al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos más a 3' del polinucleótido donante. La extensión de hibridación de 3' puede hibridarse, como máximo, con 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos más a 3' del polinucleótido donante.

Como se representa en la **Figura 30B**, la extensión de ácido nucleico crtra en el extremo 3’ del ácido nucleico dirigido a ácido nucleico puede incluir una secuencia que puede hibridarse con el extremo 5' del ADN donante. La extensión de hibridación de 3’ puede hibridarse con al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos más a 5' del polinucleótido donante. La extensión de hibridación de 3’ puede hibridarse, como máximo, con 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos más a 5' del polinucleótido donante.

La extensión de ácido nucleico crtra en el extremo 3’ del ácido nucleico dirigido a ácido nucleico puede incluir una secuencia que puede hibridarse a una región entre el extremo 3' y el extremo 5' del polinucleótido donante, como se muestra en la **Figura 30C**. La extensión de hibridación de 3’ puede hibridarse con al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos entre el extremo 3' y 5’ del polinucleótido donante. La extensión de hibridación de 3’ puede hibridarse, como máximo, con 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, nucleótidos entre el extremo 3' y 5’ del polinucleótido donante.

La extensión de ácido nucleico crtra en el extremo 3’ del ácido nucleico dirigido a ácido nucleico puede incluir una secuencia que puede hibridarse a lo largo del polinucleótido donante, como se muestra en la **Figura 30D**. El ácido nucleico dirigido a ácido nucleico puede hibridarse a lo largo de al menos aproximadamente un 20, 30, 40, 50, 60, 70, 80, 90 o 100 % del polinucleótido donante. El ácido nucleico dirigido a ácido nucleico puede hibridarse a lo largo de aproximadamente un 20, 30, 40, 50, 60, 70, 80, 90 o 100 % del polinucleótido donante. La secuencia de extensión de hibridación de 3’ puede hibridarse a lo largo de la longitud completa del polinucleótido donante con al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, emparejamientos erróneos. La secuencia de extensión de hibridación de 3’ puede hibridarse a lo largo de toda la longitud del polinucleótido donante con, como máximo, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más, emparejamientos erróneos.

La extensión de ácido nucleico crtra en el extremo 3’ del ácido nucleico dirigido a ácido nucleico (por ejemplo, extensión de hibridación de 3') puede comprender una secuencia que puede usarse como plantilla y ser convertida, por ejemplo, por una transcriptasa inversa para generar ácido nucleico híbrido (por ejemplo, el ácido nucleico resultante es un híbrido de ARN-ADN, en donde el ácido nucleico recién transcrito puede ser ADN), como se muestra en la **Figura 30E**. Los ejemplos de transcriptasas inversas incluyen SuperScript, ThermoScript, transcriptasa inversa de VIH y transcriptasa inversa de MMLV. La transcriptasa inversa puede extender la secuencia de polinucleótido donante a partir de la plantilla de extensión de hibridación de 3'.

La extensión de ácido nucleico crtra en el extremo 3' del ácido nucleico dirigido a ácido nucleico puede incorporar una secuencia de ácido nucleico que puede unirse a una proteína de unión a ARN (RBP). La proteína de unión a ARN puede fusionarse con una proteína de unión a ADN (DBP), como se muestra en la **Figura 30F**. La proteína de unión al ADN puede unirse al polinucleótido donante.

Las secuencias usadas para acercar el polinucleótido donante a una rotura bicatenaria se pueden unir al extremo 5' del ácido nucleico dirigido a ácido nucleico (por ejemplo, La extensión del espaciador). Las secuencias usadas para acercar el polinucleótido donante a una rotura bicatenaria se pueden unir tanto al extremo 5’ como al extremo 3' del ácido nucleico dirigido a ácido nucleico.

La nucleasa usada en los métodos de la divulgación (por ejemplo, Cas9) puede comprender actividad nickasa, en donde la nucleasa puede introducir roturas monocatenarias en un ácido nucleico diana. Los pares de nucleasas con actividad nickasa pueden dirigirse a regiones cercanas entre sí. Una primera nucleasa puede unirse a un primer ácido nucleico dirigido a ácido nucleico que puede interactuar con un primer polinucleótido donante. Una segunda nucleasa puede unirse a un segundo ácido nucleico dirigido a ácido nucleico que puede interactuar con un segundo polinucleótido donante. El primer y segundo polinucleótido donante se pueden diseñar para hibridarse entre sí a fin de crear un polinucleótido donante bicatenario. Se pueden llevar dos polinucleótidos donante separados a un sitio de nucleasa.

En algunas realizaciones, el polinucleótido donante puede ser monocatenario. En algunas realizaciones, el polinucleótido donante puede ser bicatenario. En algunas realizaciones, el ADN donante puede ser un minicírculo. En algunas realizaciones, el polinucleótido donante puede ser un plásmido. En algunas realizaciones, el plásmido puede estar superenrollado. En algunas realizaciones, el polinucleótido donante puede estar metilado. En algunas realizaciones, el polinucleótido donante puede no estar metilado. El polinucleótido donante puede comprender una modificación. Las modificaciones pueden incluir las descritas en el presente documento, que incluyen, aunque sin limitación, biotinilación, conjugado químico y nucleótidos sintéticos.

Métodos para clonar y expresar un vector que comprende un polipéptido dirigido a un sitio y un ácido nucleico dirigido a ácido nucleico

La divulgación proporciona métodos para clonar un ácido nucleico dirigido a ácido nucleico diseñado en un vector (por ejemplo, un vector linealizado). El método se puede realizar usando cualquiera de los polipéptidos dirigidos a un sitio, ácidos nucleicos dirigidos a ácido nucleico y complejos de polipéptidos dirigidos a un sitio y ácidos nucleicos dirigidos a ácido nucleico tales como los descritos en el presente documento.

Un usuario (por ejemplo, un científico) puede diseñar oligonucleótidos de ADN monocatenarios. Los oligonucleótidos de ADN monocatenarios pueden, cuando se hibridan entre sí, codificar una secuencia espacial para dirigirse a un ácido nucleico diana. Los oligonucleótidos de ADN monocatenarios pueden tener al menos aproximadamente 5, 10, 15, 20, 25, 30, o más, nucleótidos de longitud. Los oligonucleótidos de ADN monocatenarios pueden tener, como máximo, aproximadamente 5, 10, 15, 20, 25, 30, o más, nucleótidos de longitud. Los oligonucleótidos de ADN monocatenarios pueden tener 19-20 nucleótidos de longitud.

Un oligonucleótido de ADN monocatenario puede diseñarse de modo que pueda hibridarse con un ácido nucleico diana (por ejemplo, una secuencia adyacente a un motivo adyacente al protoespaciador, tal como el extremo 3' o 5' del motivo adyacente al protoespaciador). El oligonucleótido de ADN puede codificar una secuencia correspondiente a la cadena codificante o no codificante de la secuencia de ácido nucleico diana.

Los oligonucleótidos monocatenarios pueden comprender una primera parte que puede hibridarse y/o es complementaria a un ácido nucleico diana. Los oligonucleótidos monocatenarios pueden comprender una primera parte que puede hibridarse y/o es complementaria a otro oligonucleótido monocatenario. El oligonucleótido monocatenario puede comprender una segunda parte que puede hibridarse con una secuencia en el vector linealizado. En otras palabras, un par de oligonucleótidos monocatenarios puede comprender una primera parte que se hibrida entre sí y una segunda parte que comprende salientes monocatenarios, en donde los salientes pueden hibridarse a extremos adhesivos en el vector linealizado. En algunos casos, un saliente comprende 5'-GTTTT-3'. En algunos casos, un saliente comprende 5'-CGGTG-3'.

Los nucleótidos de ADN monocatenarios se pueden hibridar entre sí para generar un oligonucleótido bicatenario. Los nucleótidos de ADN monocatenarios se pueden hibridar entre sí en un tampón de hibridación de oligonucleótidos (por ejemplo, que comprende Tris-HCl, EDTA y NaCl). El oligonucleótido bicatenario se puede diluir a una concentración de trabajo (por ejemplo, una concentración adecuada para la unión a un plásmido linealizado). El oligonucleótido bicatenario diluido puede ligarse en un vector linealizado. La ligadura puede realizarse en un tampón de ligadura (por ejemplo, que comprende Tris-HCl, $MgCl_2$, ATP) y con una ligasa (por ejemplo, ADN ligasa de T4). El oligonucleótido bicatenario se puede ligar en un vector linealizado en una región dentro de la secuencia que codifica el ácido nucleico dirigido a ácido nucleico. En otras palabras, el vector linealizado puede linealizarse en un punto dentro de la región que codifica el ácido nucleico dirigido a ácido nucleico, en donde la linealización genera extremos adhesivos que son complementarios a los extremos adhesivos del oligonucleótido bicatenario. Cuando el oligonucleótido bicatenario se liga al vector, puede generar una secuencia que codifica un ácido nucleico dirigido a ácido nucleico diseñado que comprende una secuencia espaciadora correspondiente a la secuencia oligonucleotídica bicatenaria.

El vector ligado puede transformarse en células químicamente competentes (por ejemplo, DH5-alfa, Top10) y seleccionarse para la expresión del vector ligado correctamente (por ejemplo, mediante cribado con antibiótico). Los transformantes seleccionados pueden analizarse para detectar la presencia de una inserción mediante secuenciación. La secuenciación puede realizarse usando un cebador de secuenciación que pueda hibridarse con una parte del vector.

El vector ligado correctamente puede prepararse (por ejemplo, mediante preparación de ADN a gran escala, maxiprep) y purificarse. El vector, que comprende un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico, en donde el ácido nucleico ácido dirigido a ácido nucleico comprende los oligonucleótidos de ADN bicatenarios, se puede introducir (por ejemplo, transfectar) en una estirpe celular de elección (por ejemplo, estirpe celular de mamífero).

Ejemplos

Los ejemplos que no pertenecen a la invención son solo para fines ilustrativos.

EJEMPLO 1: MODIFICACIÓN DE UN POLIPÉPTIDO DIRIGIDO A UN SITIO PARA ALTERAR LA ESPECIFICIDAD HACIA PAM

En algunos aspectos, la divulgación proporciona un polipéptido modificado dirigido a un sitio que se modifica para alterar la especificidad hacia PAM. Un ácido nucleico que codifica un polipéptido dirigido a un sitio modificado se introduce en las células mediante transfección. El polipéptido dirigido a un sitio modificado comprende un dominio nucleasa HNH o RuvC insertado. El polipéptido modificado dirigido a un sitio comprende un parche altamente básico modificado. Un ácido nucleico dirigido a ácido nucleico que comprende un espaciador que puede hibridarse con el ácido nucleico diana también se introduce en las células mediante transfección. El polipéptido modificado dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico forman un complejo. El complejo es guiado al ácido nucleico diana por el ácido nucleico dirigido a ácido nucleico. Una vez hibridado con el ácido nucleico dirigido a ácido nucleico, el ácido nucleico diana es escindido por los dominios nucleasa del polipéptido dirigido a un sitio. En algunos aspectos, el polipéptido modificado dirigido a un sitio se une al ácido nucleico diana con una K_d inferior.

En algunos aspectos, también se introduce un polinucleótido donante en las células. En algunos casos, el polinucleótido donante, una parte de un polinucleótido donante, una copia del polinucleótido donante o una parte de una copia del polinucleótido donante se inserta en el ácido nucleico diana escindido.

EJEMPLO 2: MODIFICACIÓN DE UN POLIPÉPTIDO DIRIGIDO A UN SITIO PARA ALTERAR LA ESPECIFICIDAD HACIA EL ÁCIDO NUCLEICO DIANA

En algunos aspectos, la divulgación proporciona un polipéptido modificado dirigido a un sitio que se modifica para alterar la especificidad hacia ácido nucleico diana. Se introduce un polipéptido modificado dirigido a un sitio en células. El polipéptido modificado dirigido a un sitio comprende una modificación en el parche altamente básico y/o el dominio de tipo HNH. Un ácido nucleico dirigido a ácido nucleico que comprende un espaciador que puede hibridarse con el ácido nucleico diana también se introduce en las células. El polipéptido modificado dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico forman un complejo. El complejo es guiado al ácido nucleico diana por el ácido nucleico dirigido a ácido nucleico. Una vez hibridado con el ácido nucleico dirigido a ácido nucleico, el ácido nucleico diana es escindido por el polipéptido dirigido a un sitio. En algunos aspectos, el polipéptido modificado dirigido a un sitio se une al ácido nucleico diana con una K_d inferior.

En algunos aspectos, también se introduce un polinucleótido donante en las células. En algunos casos, el polinucleótido donante, una parte de un polinucleótido donante, una copia del polinucleótido donante o una parte de una copia del polinucleótido donante se inserta en el ácido nucleico diana escindido.

EJEMPLO 3: EXPRESIÓN RECOMBINANTE DE UN POLIPÉPTIDO DIRIGIDO A UN SITIO

Se puede ensamblar una secuencia de ADN recombinante que codifica un polipéptido modificado dirigido a un sitio, y permite la expresión del polipéptido modificado dirigido a un sitio en un organismo hospedador. La secuencia de ADN recombinante comprende una secuencia promotora, y puede comprender además un marcador de afinidad para la purificación, o un marcador de epítopo. En un ejemplo no limitante, un plásmido comprende la secuencia de ADN recombinante para la expresión del polipéptido dirigido a un sitio modificado.

Producción de proteína recombinante.

Un plásmido que codifica el polipéptido modificado dirigido a un sitio se introduce en células bacterianas (por ejemplo, *E. coli*). El polipéptido se expresa en células bacterianas y luego se purifica del lisado celular usando métodos de cromatografía. La actividad del polipéptido dirigido a un sitio modificado se mide usando métodos de ensayo diseñados para determinar la especificidad del polipéptido modificado, la secuencia de PAM, el perfil de especificidad del polipéptido dirigido a un sitio y la preferencia del ácido nucleico (por ejemplo, ADN o ARN, o ácidos nucleicos modificados.

El software está diseñado para elegir sitios que puedan cortarse usando el polipéptido modificado dirigido a un sitio. Las secuencias de ARN guía están diseñadas para dirigir la actividad del polipéptido dirigido a un sitio. Una vez diseñado, el polipéptido dirigido a un sitio se usa para escindir ácidos nucleicos.

Introducción del polipéptido modificado dirigido a un sitio en las células

El péptido modificado se introduce en las células para dirigirse a los sitios de ácido nucleico. Un polipéptido que conserva la actividad nucleasa se usa para introducir roturas de ADN monocatenarias o bicatenarias en el ADN diana. Se usa un polipéptido con actividad de unión al ADN, pero no de corte del ADN, para unir el ADN bicatenario a una célula. Esto puede usarse para efectuar la activación o represión de la transcripción.

EJEMPLO 4: SELECCIÓN DE SITIOS PARA MODIFICACIÓN DENTRO DE SECUENCIAS DE CAS9

Como se ha descrito anteriormente, los sistemas CRISPR de Tipo II que contienen ortólogos de Cas9 se pueden clasificar en tres grupos (Tipo II-A, Tipo II-B y Tipo II-C) según el análisis de sus locus de Cas de CRISPR. Los ortólogos de Cas9 dentro de estos grupos pueden definirse en términos generales por dos clados que comprenden secuencias más cortas (tipo Cas9/ Csn1) y más largas (tipo Cas9/Csx 12). Además de estos grupos más grandes, puede haber dos familias adicionales de homólogos que comprenden dominios HNH y RuvC dispuestos con una topología de tipo Cas9, pero con diferencias significativas en la longitud y secuencia de inserciones entre elementos de secuencia conservados. Las predicciones de estructura secundaria y las alineaciones de secuencia se usan para definir regiones del polipéptido para modificación. Las regiones que se encuentran entre elementos de estructura secundaria o entre regiones de alta conservación de secuencia se seleccionan como candidatos para inserciones o eliminaciones. Las regiones que tienen similitud con dominios de estructura conocida se analizan para identificar regiones específicas para insertar o eliminar secuencias.

La **Figura 35** muestra la secuencia de alineación de CDD TIGR0185 para un pequeño número de diversos ortólogos de Cas9. Los aminoácidos con una "X" debajo de ellos se consideran similares. Los aminoácidos con una "Y" debajo de ellos pueden considerarse altamente conservados o idénticos en todas las secuencias. Los restos de aminoácidos sin una "X" o una "Y" no se conservan. Esta alineación no incluye la región C-terminal (que corresponde aproximadamente a los restos de aminoácidos 1100-1350) de los ortólogos de Cas9 más largos. Las secuencias enumeradas en la **Figura 35** se enumeran en la Tabla 6.

Tabla 6. Secuencias enumeradas en la **Figura 35**.

Acceso del Genbank		Especie
gi	22533915	2603V/R de *Streptococcus agalactiae*
gi	34483507	*Wolinella succinogenes*
gi	12721472	*Pasteurella multocida* subsp. *multocida* str. Pm70
gi	24377777	UA159 de *Streptococcus mutans*
gi	13622193	MI GAS de *Streptococcus pyogenes*
gi	41815893	ATCC 35405 de *Treponema denticola*
gi	218767588	Z2491 de *Neisseria meningitidis*
gi	157150687	*Streptococcus gordonii* str. *Challis* substr. CHI
gi	294660600	*Mycoplasma gallisepticum* str. R(bajo)
gi	218563121	NCTC 11168 = ATCC 700819 de *Campylobacter jejuni* subsp. *jejuni*
gi	370792169	ATCC 33091 de *Listeria innocua*

Las ubicaciones para la inserción de nuevos dominios funcionales, regiones para la inserción de secuencias alternativas, o para la eliminación o reducción del tamaño de la región que modifican la actividad de Cas9 pueden incluir, pero sin limitación, las regiones resaltadas en la Tabla 7. Los números representan los números de secuencia de aminoácidos basados en la secuencia de Cas9 de MI GAS de *Streptococcus pyogenes*.

Tabla 7. Ubicaciones ilustrativas para modificar Cas9

gi |13622193	**MI GAS de *Streptococcus pyogenes***		
Sitio de inserción/eliminación	**inicio**	**final**	**Longitud**
1	22	42	20
2	97	134	37
3	170	312	142
4	350	400	50
5	426	444	18
6	455	492	37
7	528	541	13
8	578	589	11
9	600	612	12
10	628	634	6
11	654	658	4
12	684	692	8
13	710	727	17
14	753	756	3
15	792	794	2
16	801	804	3
17	826	834	8
18	873	881	8
19	893	915	22
20	939	952	13
21	971	974	3
22	1005	1029	24
23	1096	1225	129
24	1105	1138	33

Una vez que se ha identificado una región como una ubicación potencial para insertar una nueva secuencia de polipéptido en la proteína, o eliminar una región de la proteína, la secuencia de ADN que codifica la proteína se modifica para incorporar la modificación.

EJEMPLO 5: ENRIQUECIMIENTO DE SECUENCIA DEL ÁCIDO NUCLEICO DIANA UNIDO AL POLIPEPTIDO DIRIGIDO A UN SITIO

La divulgación proporciona métodos para el enriquecimiento de secuencias sin amplificación usando polipéptidos dirigidos a un sitio.

En algunos aspectos, el método comprenderá a) poner en contacto un ácido nucleico diana con un complejo que comprende un ácido nucleico dirigido a ácido nucleico y un polipéptido dirigido a un sitio; b) escindir el ácido nucleico diana; c) purificar el ácido nucleico diana; y d) secuenciar el ácido nucleico diana, en donde dicho ácido nucleico diana es enriquecido.

En algunos aspectos, el polipéptido dirigido a un sitio estará enzimáticamente inactivo. El uso de un polipéptido dirigido a un sitio enzimáticamente inactivo facilitará la unión del ácido nucleico diana al complejo del polipéptido dirigido a un sitio. En algunos aspectos, el polipéptido dirigido a un sitio será enzimáticamente activo.

En algunos aspectos, el enriquecimiento de secuencia se realiza fuera de las células (por ejemplo, muestra libre de células). Por ejemplo, una muestra comprenderá ADN genómico purificado. En algunos aspectos, el enriquecimiento de secuencia se realizará en una muestra celular (por ejemplo, células, lisado celular).

En algunos casos, los complejos de polipéptido dirigido a un sitio-ácido nucleico dirigido a ácido nucleico se fijarán o entrecruzarán para formar complejos. Si el método se realiza en células, las células se lisarán. Las condiciones de lisis se elegirán para mantener intactos los complejos de proteína-ADN.

La muestra de ácido nucleico se tratará para fragmentar el ácido nucleico diana antes de la purificación por afinidad. La fragmentación se puede realizar a través de métodos físicos, mecánicos o enzimáticos. La fragmentación física incluirá la exposición de un polinucleótido diana al calor o a la luz ultravioleta (UV). La disrupción mecánica se usará para cortar mecánicamente un polinucleótido diana en fragmentos del intervalo deseado. La cizalla mecánica se realizará a través de varios métodos, que incluyen pipeteo repetitivo del polinucleótido diana, sonicación y nebulización. Los ácidos nucleicos diana también se fragmentarán usando métodos enzimáticos. En algunos casos, se realizará la digestión enzimática usando enzimas tales como las enzimas de restricción. Las enzimas de restricción se usarán para realizar una fragmentación específica o no específica de polinucleótidos diana. Los métodos usarán uno o más tipos de enzimas de restricción, descritas en general como enzimas de Tipo I, enzimas de Tipo II y/o enzimas de Tipo III. Las enzimas de Tipo II y de Tipo III reconocen secuencias específicas de nucleótidos dentro de una secuencia polinucleotídica bicatenaria (una "secuencia de reconocimiento" o "sitio de reconocimiento"). Tras la unión y el reconocimiento de estas secuencias, las enzimas de Tipo II y Tipo III escinden la secuencia de polinucleótidos. En algunos casos, la escisión dará lugar a un fragmento de polinucleótido con una parte de ácido nucleico monocatenario saliente, denominada "extremo adhesivo". En otros casos, la escisión no dará lugar a un fragmento con un saliente, creando un "extremo romo". Los métodos pueden comprender el uso de enzimas de restricción que generen extremos adhesivos o extremos romos.

Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio se purificarán mediante incubación con un soporte sólido. Por ejemplo, si el polipéptido dirigido a un sitio comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al marcador de biotina.

En un aspecto alternativo, Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, el ácido nucleico diana y/o el ácido nucleico dirigido a ácido nucleico, se purificarán mediante incubación con un agente de captura. El agente de captura se unirá al marcador de afinidad fusionado con el polipéptido dirigido a un sitio. El agente de captura comprenderá un anticuerpo. Por ejemplo, si el marcador de afinidad fusionado con el polipéptido dirigido a un sitio es un marcador FLAG, entonces el agente de captura será un anticuerpo anti-marcador FLAG.

En algunos aspectos, el agente de captura se purificará con un soporte sólido. Por ejemplo, si el agente de captura comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprenderá un marcador de afinidad. El marcador de afinidad comprenderá una secuencia que pueda unirse a una endorribonucleasa. En algunos casos, el marcador de afinidad comprenderá una secuencia que pueda unirse a una endorribonucleasa condicionalmente inactiva enzimáticamente. La endorribonucleasa condicionalmente enzimáticamente inactiva se unirá al, pero no escindirá, el marcador de afinidad.

En algunos aspectos, la endorribonucleasa y/o la endorribonucleasa condicionalmente enzimáticamente inactiva comprenderán un marcador de afinidad.

La endorribonucleasa condicionalmente enzimáticamente inactiva se purificará con un soporte sólido. El soporte sólido se unirá al marcador de afinidad de la endorribonucleasa condicionalmente enzimáticamente inactiva. Por ejemplo, si la endorribonucleasa condicionalmente enzimáticamente inactiva comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos, la endorribonucleasa condicionalmente enzimáticamente inactiva se inmovilizará en cualquiera de varios soportes insolubles.

En algunos aspectos del método, se realizarán dos series de purificación. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura.

En algunos aspectos, los métodos de la divulgación se usarán para el enriquecimiento de múltiples secuencias. En este aspecto, se puede poner en contacto una pluralidad de ácidos nucleicos dirigidos a ácido nucleico con una muestra de ácido nucleico, en donde cada ácido nucleico dirigido a ácido nucleico está diseñado para dirigirse un ácido nucleico diana diferente (por ejemplo, una secuencia de un genoma) dentro de la muestra de ácido nucleico.

Los complejos capturados comprenderán un ácido nucleico diana. El ácido nucleico diana se eluirá del complejo del polipéptido dirigido a un sitio mediante métodos convencionales que incluyen el lavado con alto contenido de sal, precipitación en etanol, ebullición, purificación en gel y similares.

El ADN eluido se preparará para el análisis de secuenciación mediante ligadura de uno o más adaptadores.

Los bancos de secuenciación se secuenciarán como se describe en el presente documento. Los bancos secuenciados se analizarán para identificar los polimorfismos, diagnosticar una enfermedad, determinar un curso de tratamiento para una enfermedad y/o generar bancos de anticuerpos.

EJEMPLO 6: ENRIQUECIMIENTO DE SECUENCIA DEL ÁCIDO NUCLEICO DIANA NO UNIDO A UN COMPLEJO QUE COMPRENDE UN POLIPÉPTIDO DIRIGIDO A UN SITIO

En algunos aspectos, el enriquecimiento de secuencia se realizará con un polipéptido dirigido enzimáticamente a un sitio. En algunos casos, el polipéptido dirigido a un sitio será enzimáticamente activo. En este caso, el ácido nucleico diana no se unirá a un polipéptido dirigido a un sitio, sino que se escindirá.

Se identificará un ácido nucleico diana, y los ácidos nucleicos dirigidos a ácido nucleico se diseñarán para dirigir el polipéptido dirigido a un sitio a secuencias que flanqueen el ácido nucleico diana. La muestra se incubará con un complejo que comprenda un ácido nucleico dirigido a ácido nucleico diseñado y el polipéptido dirigido a un sitio, de modo que el polipéptido dirigido a un sitio escindirá el ADN en ambos extremos del ácido nucleico diana. Tras la escisión del ácido nucleico diana, el ácido nucleico diana se escindirá del ácido nucleico original. Se purificará el ácido nucleico diana escindido (por ejemplo, mediante electroforesis en gel, la elución de tamaño selectivo a partir de perlas, u otras perlas derivatizadas de carboxilato, o mediante la precipitación con concentraciones apropiadas de sal y PEG para precipitar preferentemente el ADN más grande o más pequeño).

En algunos aspectos, el enriquecimiento de secuencia se realiza fuera de las células (por ejemplo, muestra libre de células). Por ejemplo, una muestra comprenderá ADN genómico purificado. En algunos aspectos, el enriquecimiento de secuencia se realizará en una muestra celular (por ejemplo, células, lisado celular).

Si el método se realiza en células, las células se lisarán. Las condiciones de lisis se elegirán para mantener intactos los complejos de proteína-ADN.

En algunos aspectos, el ácido nucleico diana que se va a secuenciar no estará unido a un ácido nucleico dirigido a ácido nucleico ni/o a un polipéptido dirigido a un sitio. En este aspecto, el ácido nucleico unido al polipéptido dirigido a un sitio y/o al ácido nucleico dirigido a ácido nucleico se purificará. La purificación del polipéptido dirigido a un sitio procederá como se describe anteriormente en el presente documento. Brevemente, los complejos que comprenden el polipéptido dirigido a un sitio se purificarán mediante incubación con un soporte sólido. Por ejemplo, si el polipéptido dirigido a un sitio comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al marcador de biotina.

En un aspecto alternativo, Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, ácido nucleico dirigido a ácido nucleico y ácido nucleico no diana, se purificarán mediante incubación con un agente de captura. El agente de captura se unirá al marcador de afinidad fusionado con el polipéptido dirigido a un sitio. El agente de captura comprenderá un anticuerpo. Por ejemplo, si el marcador de afinidad fusionado con el polipéptido dirigido a un sitio es un marcador FLAG, entonces el agente de captura será un anticuerpo anti-marcador FLAG.

El agente de captura se purificará con un soporte sólido. Por ejemplo, si el agente de captura comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos, los métodos de la divulgación se usarán para el enriquecimiento de múltiples secuencias. En este aspecto, se puede introducir una pluralidad de ácidos nucleicos dirigidos a ácido nucleico en una célula, en donde cada ácido nucleico dirigido a ácido nucleico está diseñado para dirigirse un ácido nucleico diana diferente (por ejemplo, una secuencia de un genoma).

El complejo capturado no comprenderá un ácido nucleico diana.

El ácido nucleico diana comprenderá el ácido nucleico que no está unido a los complejos que comprenden el polipéptido dirigido a un sitio. El ácido nucleico diana puede recogerse mediante métodos patrón de purificación de ácido nucleico (por ejemplo, un kit de purificación por PCR disponible en el mercado, un gel de agarosa).

El ácido nucleico diana recogido se preparará para el análisis de secuenciación (por ejemplo, secuenciación profunda) mediante ligadura de uno o más adaptadores como se describe en el presente documento.

El ácido nucleico diana secuenciado se analizará para identificar polimorfismos, diagnosticar una enfermedad, determinar un curso de tratamiento para una enfermedad y/o generar bancos de anticuerpos.

EJEMPLO 7: SECUENCIACIÓN

Los ácidos nucleicos diana eluidos se prepararán para el análisis de secuenciación. La preparación para el análisis de secuenciación incluirá la generación de bancos de secuenciación del ácido nucleico diana eluido. El análisis de secuenciación determinará la identidad y frecuencia de los sitios de unión fuera de la diana de los polipéptidos dirigidos a un sitio.

La determinación de las secuencias se realizará usando métodos que determinen muchas secuencias de ácido nucleico (normalmente, de miles a miles de millones) de una manera intrínsecamente paralela, donde se leen muchas secuencias preferentemente en paralelo mediante un proceso en serie de alto rendimiento. Dichos métodos pueden incluir, pero sin limitación, pirosecuenciación (por ejemplo, como la comercializada por 454 Life Sciences, Inc., Branford, Conn.); secuenciación por ligadura (por ejemplo, como la comercializada en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, Calif.); La secuenciación por síntesis usando nucleótidos modificados (tales como los comercializados en la tecnología TruSeq™ y HiSeq™ por Illumina, Inc., San Diego, Calif., HeliScope™ de Helicos Biosciences Corporation, Cambridge, Mass. yPacBio RS de Pacific Bio sciences of California, Inc., Menlo Park, Calif.), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, Calif.); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, Calif.); tecnologías de secuenciación basadas en nanoporos (por ejemplo, las desarrolladas por Oxford Nanopore Technologies, LTD, Oxford, Reino Unido), secuenciación capilar (por ejemplo, comercializada en MegaBACE por Molecular Dynamics), secuenciación electrónica, secuenciación de moléculas individuales (por ejemplo, tal como la comercializada en tecnología SMRT™ de Pacific Biosciences, Menlo Park, Calif.), secuenciación microfluídica de gotas, secuenciación por hibridación (tal como la comercializada por Affymetrix, Santa Clara, California), secuenciación de bisulfato y otros métodos de secuenciación altamente paralelizados conocidos.

En algunos aspectos, la secuenciación se realizará mediante análisis de micromatrices.

EJEMPLO 8: GENERACIÓN DE BANCOS ANTICUERPOS

Los métodos descritos en el presente documento se usarán para generar bancos de proteínas (por ejemplo, bancos de anticuerpos). Los bancos de proteínas serán útiles para preparar bancos de expresión, que se usarán para seleccionar proteínas (por ejemplo, anticuerpos) para uso en agentes terapéuticos, reactivos y/o de diagnóstico. Los bancos de proteínas también serán útiles para sintetizar y/o clonar anticuerpos adicionales.

Los bancos de proteínas se generarán mediante el diseño de un ácido nucleico dirigido a ácido nucleico para hibridarse con secuencias de ácido nucleico diana que codifican inmunoglobulinas. Los complejos que comprenden un polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico se purificarán usando los métodos descritos en el presente documento. En algunos aspectos, el ácido nucleico que se hibrida con el ácido nucleico dirigido a ácido nucleico será el ácido nucleico diana, y se eluirá y secuenciará, usando los métodos que se describen en el presente documento. En algunos aspectos, el ácido nucleico que se hibrida con el ácido nucleico dirigido a ácido nucleico no será el ácido nucleico diana. El ácido nucleico diana será el ácido nucleico que se escinda entre los sitios de escisión de una pluralidad de complejos (por ejemplo, complejos que comprenden un polipéptido dirigido a un sitio y ácido nucleico dirigido a ácido nucleico). El ácido nucleico diana escindido se purificará y secuenciará, usando los métodos que se describen en el presente documento.

EJEMPLO 9: GENOTIPADO

Los métodos descritos en el presente documento se usarán para realizar la tipificación del antígeno leucocitario humano (HLA). Los genes HLA son algunos de los genes más polimórficos de los seres humanos. Comprender los genotipos de estas regiones será importante para obtener una buena coincidencia para los trasplantes de tejidos y

órganos.

Para realizar la tipificación de HLA, se diseñará un ácido nucleico dirigido a ácido nucleico para hibridarse con secuencias de ácido nucleico diana en genes HLA. Los complejos que comprenden un polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico se purificarán usando los métodos descritos en el presente documento. En algunos aspectos, el ácido nucleico que se hibrida con el ácido nucleico dirigido a ácido nucleico será el ácido nucleico diana, y se eluirá y secuenciará, usando los métodos que se describen en el presente documento. En algunos aspectos, el ácido nucleico que se hibrida con el ácido nucleico dirigido a ácido nucleico no será el ácido nucleico diana. El ácido nucleico diana será el ácido nucleico que se escinda entre los sitios de escisión de una pluralidad de complejos (por ejemplo, complejos que comprenden un polipéptido dirigido a un sitio y ácido nucleico dirigido a ácido nucleico). El ácido nucleico diana escindido se purificará y secuenciará, usando los métodos que se describen en el presente documento.

EJEMPLO 10: INMUNOPRECIPITACIÓN DE POLIPÉPTIDO DIRIGIDO A UN SITIO

La divulgación proporciona métodos para la inmunoprecipitación y secuenciación de nucleasas (NIP-Seq). En algunos aspectos, el método comprenderá a) poner en contacto una muestra de ácido nucleico con un polipéptido dirigido a un sitio enzimáticamente inactivo, en donde polipéptido dirigido a un o enzimáticamente inactivo se une a un ácido nucleico diana, formando así un complejo; b) capturar el complejo con un agente de captura; y c) secuenciar el ácido nucleico diana. En algunos aspectos, el método comprenderá además d) determinar la identidad del sitio de unión fuera de la diana.

En algunos aspectos, los métodos de la divulgación se realizarán fuera de las células. Por ejemplo, una muestra comprenderá ADN genómico purificado.

Los complejos de polipéptido dirigido a un sitio-ácido nucleico dirigido a ácido nucleico se fijarán o entrecruzarán para formar complejos.

El ácido nucleico (por ejemplo, ADN genómico) se tratará para fragmentar el ADN antes de la purificación por afinidad. La fragmentación se puede realizar a través de métodos físicos, mecánicos o enzimáticos. La fragmentación física puede incluir la exposición de un polinucleótido diana al calor o a la luz ultravioleta (UV). La disrupción mecánica puede usarse para cortar mecánicamente un polinucleótido diana en fragmentos del intervalo deseado. La cizalla mecánica se puede lograr a través de varios métodos conocidos en la técnica, que incluyen pipeteo repetitivo del polinucleótido diana, sonicación y nebulización. Los polinucleótidos diana también pueden fragmentarse usando métodos enzimáticos. En algunos casos, se puede realizar la digestión enzimática usando enzimas tales como las enzimas de restricción. Las enzimas de restricción pueden usarse para realizar una fragmentación específica o no específica de polinucleótidos diana. Los métodos pueden usar uno o más tipos de enzimas de restricción, descritas en general como enzimas de Tipo I, enzimas de Tipo II y/o enzimas de Tipo III. Las enzimas de Tipo II y Tipo III, en general, están disponibles en el mercado, y son bien conocidas en la técnica. Las enzimas de Tipo II y de Tipo III reconocen secuencias específicas de nucleótidos nucleotídicos dentro de una secuencia polinucleotídica bicatenaria (una "secuencia de reconocimiento" o "sitio de reconocimiento"). Tras la unión y el reconocimiento de estas secuencias, las enzimas de Tipo II y Tipo III escinden la secuencia de polinucleótidos. En algunos casos, la escisión dará lugar a un fragmento de polinucleótido con una parte de ADN monocatenario saliente, denominada "extremo adhesivo". En otros casos, la escisión no dará lugar a un fragmento con un saliente, creando un "extremo romo". Los métodos pueden comprender el uso de enzimas de restricción que generen extremos adhesivos o extremos romos.

Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio se purificarán mediante incubación con un soporte sólido. Por ejemplo, si el polipéptido dirigido a un sitio comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al marcador de biotina.

En un aspecto alternativo, Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, el ácido nucleico diana y/o el ácido nucleico dirigido a ácido nucleico, se purificarán mediante incubación con un agente de captura. El agente de captura se unirá al marcador de afinidad fusionado con el polipéptido dirigido a un sitio. El agente de captura comprenderá un anticuerpo. Por ejemplo, si el marcador de afinidad fusionado con el polipéptido dirigido a un sitio es un marcador FLAG, entonces el agente de captura será un anticuerpo anti-marcador FLAG.

El agente de captura se purificará con un soporte sólido. Por ejemplo, si el agente de captura comprende un marcador de biotina, la perla se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos del método, se realizarán dos o más series de purificación. Una primera serie de purificación comprenderá la purificación con un soporte sólido que se puede unir al marcador de afinidad del agente de captura, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se puede unir al marcador de afinidad del polipéptido dirigido a un sitio. Una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura.

En algunos aspectos, el método se usará para optimizar la especificidad de unión de un polipéptido dirigido a un sitio realizando el método más de una vez.

El complejo capturado comprenderá un polipéptido dirigido a un sitio y un ácido nucleico diana. El ácido nucleico diana se eluirá del complejo del polipéptido dirigido a un sitio mediante métodos convencionales que incluyen el lavado con alto contenido de sal, precipitación en etanol, ebullición, purificación en gel y similares.

El ADN eluido se preparará para el análisis de secuenciación usando métodos patrón. Los bancos de secuenciación se secuenciarán y analizarán para identificar la secuencia y la frecuencia de los sitios de unión a nucleasas.

En algunos aspectos, el método se realizará una pluralidad de veces. En algunos aspectos, el método comprende además recopilar datos y almacenar datos. Los datos pueden almacenarse recopilados y almacenados en un servidor informático.

EJEMPLO 11: INMUNOPRECIPTIACIÓN IN VIVO DE POLIPÉPTIDO DIRIGIDO A UN SITIO

En algunos aspectos, el método comprenderá: a) introducir un polipéptido dirigido a un sitio enzimáticamente inactivo en una célula, en donde polipéptido dirigido a un o enzimáticamente inactivo se une a un ácido nucleico diana, formando así un complejo; b) capturar el complejo con un agente de captura; y c) secuenciar el ácido nucleico diana. En algunos aspectos, el método comprenderá además d) determinar la identidad del sitio de unión fuera de la diana.

En algunos casos, el polipéptido dirigido a un sitio comprenderá un marcador de afinidad. Los polipéptidos que comprenden un marcador de afinidad se han descrito en el presente documento.

Las células se fijarán o entrecruzarán. Las células fijadas y/o entrecruzadas se lisarán. Las condiciones de lisis se elegirán para mantener intactos los complejos de proteína-ADN. El lisado se tratará para fragmentar el ADN antes de la purificación por afinidad. Las técnicas de fragmentación adecuadas se describen en el presente documento.

Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, el ácido nucleico diana y/o el ácido nucleico dirigido a ácido nucleico, se purificarán del lisado mediante incubación con un soporte sólido. Por ejemplo, si el polipéptido dirigido a un sitio comprende un marcador de biotina, el soporte sólido se recubrirá con avidina o estreptavidina para unirse al marcador de biotina.

En un aspecto alternativo, Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, se purificarán del lisado mediante incubación con un agente de captura. El agente de captura se unirá al marcador de afinidad fusionado con el polipéptido dirigido a un sitio. El agente de captura comprenderá un anticuerpo. Por ejemplo, si el marcador de afinidad fusionado con el polipéptido dirigido a un sitio es un marcador FLAG, entonces el agente de captura será un anticuerpo anti-marcador FLAG.

En algunos aspectos, el agente de captura comprenderá un marcador de afinidad. El agente de captura se purificará con un soporte sólido. El soporte sólido se unirá al marcador de afinidad del agente de captura. Por ejemplo, si el agente de captura comprende un marcador de biotina, la perla se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos del método, se realizarán dos series de purificación. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del agente de captura.

En algunos aspectos, el método se usará para optimizar la especificidad de unión de un polipéptido dirigido a un sitio realizando el método más de una vez.

El complejo capturado comprenderá un polipéptido dirigido a un sitio y un ácido nucleico diana. El ácido nucleico diana se eluirá del complejo del polipéptido dirigido a un sitio mediante métodos convencionales que incluyen el lavado con alto contenido de sal, precipitación en etanol, ebullición, purificación en gel y similares.

El ADN eluido se preparará para el análisis de secuenciación. Los bancos de secuenciación se crearán a partir del ácido nucleico diana eluido. Los bancos de secuenciación se secuenciarán y analizarán para identificar la secuencia y la frecuencia de los sitios de unión a nucleasas.

EJEMPLO 12: INMUNOPRECIPITACIÓN CON UN AGENTE DE CAPTURA DE ENDORRIBONULEASA ENZIMÁTICAMENTE INACTIVO

Un método para determinar la identidad de un sitio de unión fuera de la diana de una nucleasa comprenderá: a) poner en contacto una muestra de ácido nucleico con un polipéptido dirigido a un sitio enzimáticamente inactivo y un ácido nucleico ácido dirigido a ácido nucleico, en donde el polipéptido dirigido a un sitio enzimáticamente inactivo y el ácido nucleico ácido dirigido a ácido nucleico se unen a un ácido nucleico diana, formando así un complejo; b) capturar el complejo con un agente de captura, en donde el agente de captura comprende un polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo; c) secuenciar el ácido nucleico diana; y d) determinar la identidad del sitio de unión fuera de la diana. Este método está diseñado para realizarse en una muestra de ácido nucleico libre de células, y/o una muestra de ácido nucleico procedente de una célula.

Los complejos de polipéptido dirigido a un sitio-ácido nucleico dirigido a ácido nucleico se fijarán o entrecruzarán para formar complejos.

Si la muestra de ácido nucleico procede de células, los complejos fijos y/o entrecruzados se lisarán. Las condiciones de lisis se pueden escoger para mantener intactos los complejos de proteína-ADN. El lisado se tratará para fragmentar el ADN antes de la purificación por afinidad. Si la muestra de ácido nucleico procede de una muestra libre de células, el ácido nucleico libre de células se tratará para fragmentar el ADN. Las técnicas de fragmentación adecuadas se describen en el presente documento.

En algunos aspectos, el ácido nucleico dirigido a ácido nucleico comprenderá un marcador de afinidad. El marcador de afinidad comprenderá una secuencia que puede unirse a un polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo. En algunos casos, el marcador de afinidad comprenderá una secuencia que pueda unirse a una endorribonucleasa condicionalmente inactiva enzimáticamente. En algunos casos, el marcador de afinidad comprenderá una secuencia que pueda unirse a una proteína Csy4 condicionalmente inactiva enzimáticamente. El polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo se unirá, pero no escindirá, el marcador de afinidad. El marcador de afinidad comprenderá la secuencia de nucleótidos 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3'. El marcador de afinidad se introducirá en un ácido nucleico usando métodos recombinantes convencionales.

Una vez fragmentados, los complejos que comprenden el polipéptido dirigido a un sitio, el ácido nucleico diana y/o el ácido nucleico dirigido a ácido nucleico se purificarán mediante incubación con un polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo (por ejemplo, la variante Csy4).

En algunos aspectos, el polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo comprenderá un marcador de afinidad.

El polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo se purificará con un soporte sólido. El soporte sólido se unirá al marcador de afinidad del polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo. Por ejemplo, si el polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo comprende un marcador de biotina, la perla se recubrirá con avidina o estreptavidina para unirse al agente de captura biotinilado.

En algunos aspectos, el polipéptido dirigido a un sitio enzimáticamente inactivo se inmovilizará en cualquiera de varios soportes insolubles. En algunos aspectos del método, se realizarán dos series de purificación.

En algunos aspectos del método, se realizarán dos series de purificación. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo (por ejemplo, variante Csy4), y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptidos dirigido a un sitio. En algunos casos, una primera serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio, y una segunda serie de purificación comprenderá la purificación con un soporte sólido que se unirá al marcador de afinidad del polipéptido dirigido a un sitio condicionalmente enzimáticamente inactivo (por ejemplo, variante Csy4).

En algunos aspectos, el método se usará para optimizar la especificidad de unión de un polipéptido dirigido a un sitio realizando el método más de una vez.

El complejo capturado comprenderá un polipéptido dirigido a un sitio y un ácido nucleico diana. El ácido nucleico diana se eluirá del complejo del polipéptido dirigido a un sitio mediante métodos convencionales que incluyen el lavado con alto contenido de sal, precipitación en etanol, ebullición, purificación en gel y similares.

El ADN eluido se preparará para el análisis de secuenciación. Los bancos de secuenciación se secuenciarán y analizarán para identificar la secuencia y la frecuencia de los sitios de unión a nucleasas.

EJEMPLO 13: SECUENCIACIÓN

Los ácidos nucleicos diana eluidos se prepararán para el análisis de secuenciación. La preparación para el análisis de

secuenciación incluirá la generación de bancos de secuenciación del ácido nucleico diana eluido. El análisis de secuenciación determinará la identidad y frecuencia de los sitios de unión fuera de la diana de los polipéptidos dirigidos a un sitio.

La determinación de las secuencias también se realizará usando métodos que determinen muchas secuencias de ácido nucleico (normalmente, de miles a miles de millones) de una manera intrínsecamente paralela, donde se leen muchas secuencias preferentemente en paralelo mediante un proceso en serie de alto rendimiento. Dichos métodos incluyen, pero sin limitación, pirosecuenciación (por ejemplo, como la comercializada por 454 Life Sciences, Inc., Branford, Conn.); secuenciación por ligadura (por ejemplo, como la comercializada en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, Calif.); La secuenciación por síntesis usando nucleótidos modificados (tales como los comercializados en la tecnología TruSeq™ y HiSeq™ por Illumina, Inc., San Diego, Calif., HeliScope™ de Helicos Biosciences Corporation, Cambridge, Mass. y PacBio RS de Pacific Biosciences of California, Inc., Menlo Park, Calif.), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, Calif.); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, Calif.); tecnologías de secuenciación basadas en nanoporos (por ejemplo, las desarrolladas por Oxford Nanopore Technologies, LTD, Oxford, Reino Unido) y otros métodos de secuenciación altamente paralelizados conocidos.

EJEMPLO 14: Modificación de un ácido nucleico diana con una proteína efectora

Un vector que comprende un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y/o una proteína efectora se introduce en una célula. Una vez dentro de la célula, se forma un complejo que comprende los elementos codificados en el vector. El ácido nucleico dirigido a ácido nucleico se modifica con una secuencia de unión a la proteína Csy4. La proteína efectora, Csy4, se une al ácido nucleico dirigido a ácido nucleico modificado. Csy4 comprende una secuencia no nativa (por ejemplo, una fusión), que modifica un ácido nucleico diana. La secuencia no nativa es una secuencia que modifica la transcripción del ácido nucleico diana. La secuencia no nativa es un factor de transcripción. El factor de transcripción aumenta el nivel de transcripción del ácido nucleico diana. En algunos casos, la secuencia no nativa es una metilasa. La metilasa produce aumentos en la metilación del ácido nucleico diana. En algunos casos, la secuencia no nativa es una desmetilasa. La desmetilasa produce una disminución en la metilación del ácido nucleico diana. En algunos casos, la secuencia no nativa es un péptido de reclutamiento de Rad51. El péptido de reclutamiento de Rad51 aumenta el nivel de recombinación homóloga en el sitio diana. En algunos casos, la secuencia no nativa es un péptido de reclutamiento de BCRA-2. El péptido de reclutamiento de BRCA-2 aumenta el nivel de recombinación homóloga en el sitio diana.

EJEMPLO 15: Uso de un polipéptido dirigido a un sitio como biosensor para un evento de movilidad genética

Uno o varios vectores que comprenden un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y/o una proteína efectora se introduce en una célula. El polipéptido dirigido a un sitio y las proteínas efectoras se fusionan con las secuencias de ubicación celular (por ejemplo, una señal de localización nuclear). Una vez dentro de la célula, se forma un complejo que comprende los elementos codificados en el/los vector/es. En algunos casos, se introducen dos vectores en la célula. El/los vector/es codifica/n una primera proteína efectora (Csy4) que comprende una primera parte inactiva de una proteína fluorescente verde (GFP) dividida y se une a un primer ácido nucleico dirigido a ácido nucleico y una segunda proteína efectora (Csy4, Cas5 o Cas6) que comprende una segunda parte inactiva de la GFP dividida y se une a un segundo ácido nucleico dirigido a ácido nucleico. El primer ácido nucleico dirigido a ácido nucleico se modifica con una primera secuencia de unión a la proteína Csy4, Cas5 o Cas6 que puede unirse por una primera proteína Csy4, Cas5 o Cas6. El segundo ácido nucleico dirigido a ácido nucleico se modifica con una primera secuencia de unión a la proteína Csy4, Cas5 o Cas6 que puede unirse por una segunda proteína Csy4, Cas5 o Cas6. En algunos aspectos, la primera proteína Csy4, Cas5 o Cas6 interactúa con la primera secuencia de unión a la proteína Csy4, Cas5 o Cas6, y la segunda proteína Csy4, Cas5 o Cas6 interactúa con la segunda secuencia de unión a la proteína Csy4, Cas5 o Cas6. Cuando el primer y el segundo ácido nucleico que se dirigen a ácido nucleico dirigen el polipéptido dirigido a un sitio para que se una a dos secuencias que están muy próximas, la primera proteína efectora y la segunda proteína efectora pondrán en contacto la primera parte inactiva de la GFP dividida con la segunda parte inactiva de la GFP dividida, para generar una GFP activa. Los ácidos nucleicos dirigidos a ácido nucleico del complejo están diseñados de manera que un ácido nucleico dirigido a ácido nucleico guía el complejo a, por ejemplo, una región en o cerca del gen Bcr, y otro ácido nucleico dirigido a ácido nucleico guía al complejo a, por ejemplo, una región en o cerca del gen Abl. Si no se ha producido un evento de translocación, el gen Bcr está en el cromosoma 22 y el gen Abl está en el cromosoma 9, y las secuencias de ácido nucleico diana están lo suficientemente separadas como para que las dos partes inactivas del sistema GFP dividido no puedan interactuar, no generando, por tanto, una señal. Si se ha producido un evento de translocación, el gen Bcr y el gen Abl se translocan de manera que los genes estén muy juntos. En este caso, las secuencias de ácido nucleico diana están lo suficientemente juntas como para que las dos partes inactivas del sistema GFP dividido se unan para formar una GFP activa. Una señal GFP puede ser detectada por un fluorómetro. La señal es indicativa de un genotipo particular resultante del evento de movilidad genética.

EJEMPLO 16: Uso de un polipéptido dirigido a un sitio como biosensor para un evento de mutación genética

El sistema descrito en el Ejemplo 15 también puede usarse para detectar la presencia de una mutación específica

dentro de una célula. En este ejemplo, se elige un primer ácido nucleico dirigido a ácido nucleico para dirigir el polinucleótido dirigido a un sitio a una secuencia nativa ubicada cerca de un sitio de mutación. El segundo ácido nucleico dirigido a ácido nucleico se elige para reconocer una secuencia mutante (por ejemplo, la secuencia mutante que se ha identificado mediante secuenciación de ADN). El ácido nucleico dirigido a ácido nucleico se elige de manera que la secuencia mutante se produzca dentro de los 12 primeros ácidos nucleicos inmediatamente 5' de la secuencia PAM en el sitio. En este caso, las secuencias de ácido nucleico diana están lo suficientemente juntas como para que las dos partes inactivas del sistema GFP dividido se unan para formar una GFP activa. Una señal GFP puede ser detectada por un fluorómetro. ro.La señal es indicativa de un genotipo particular.

EJEMPLO 17: Uso de un polipéptido dirigido a un sitio como agente terapéutico para enfermedades que comprenden un evento de movilidad genética

Uno o varios vectores que comprenden un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y/o una proteína efectora, un ácido nucleico que comprende un péptido inductor de lisis celular (por ejemplo, proteína de muerte de Adenovirus) unido operativamente a un primer promotor también se introducirán en la célula. Una vez dentro de la célula, se forma un complejo que comprende los elementos codificados en el/los vector/es. En algunos casos, se introducen dos vectores en la célula. El/los vector/es codifica/n una primera proteína efectora (que comprende una primera secuencia de proteína Csy4, Cas5 o Cas6) que se une a un primer ácido nucleico dirigido a ácido nucleico y comprende un dominio activador para un primer factor de transcripción que se une al primer promotor y una segunda proteína efectora (que comprende una segunda secuencia de proteína Csy4, Cas5 o Cas6) que se une a un segundo ácido nucleico dirigido a ácido nucleico y comprende el dominio de unión al ADN para el primer factor de transcripción. El primer ácido nucleico dirigido a ácido nucleico se modifica con una primera secuencia de unión a la proteína Csy4, Cas5 o Cas6 que puede unirse por una primera secuencia de proteína Csy4, Cas5 o Cas6. El segundo ácido nucleico dirigido a ácido nucleico se modifica con una primera secuencia de unión a la proteína Csy4, Cas5 o Cas6 que puede unirse por una segunda proteína Csy4, Cas5 o Cas6. En algunos aspectos, la primera proteína Csy4, Cas5 o Cas6 interactúa preferentemente con la primera secuencia de unión a la proteína Csy4, Cas5 o Cas6, y la segunda proteína Csy4, Cas5 o Cas6 interactúa preferentemente con la segunda secuencia de unión a la proteína Csy4, Cas5 o Cas6. Si una célula enferma comprende un genoma que contiene un evento de movilidad genética, cuando el primer y el segundo ácido nucleico que se dirigen a ácido nucleico dirigen el polipéptido dirigido a un sitio para que se una a dos secuencias que están muy próximas, la primera proteína efectora y la segunda proteína efectora acercarán el dominio activador y el dominio de unión al ADN del primer factor de transcripción. El dominio de unión al ADN del primer factor de transcripción puede unirse al primer promotor unido operativamente al péptido inductor de la lisis celular, y el dominio activador proximal inducirá la transcripción del ARN que codifica el péptido inductor de la lisis celular. En una célula no enferma, que no comprende el evento movilidad genética, el dominio de unión a ADN y los dominios de activador del primer factor de transcripción no se acercan estrechamente, y no habrá ninguna transcripción del péptido inductor de la lisis celular.

Los ácidos nucleicos dirigidos a ácido nucleico del complejo están diseñados de manera que un ácido nucleico dirigido a ácido nucleico guía el complejo a, por ejemplo, una región en o cerca del gen Bcr, y otro ácido nucleico dirigido a ácido nucleico guía al complejo a, por ejemplo, una región en o cerca del gen Abl. En una célula no enferma, no se ha producido un evento de translocación, el gen Bcr está en el cromosoma 22 y el gen Abl está en el cromosoma 9, y las secuencias de ácido nucleico diana están lo suficientemente separadas como para que las dos partes inactivas del sistema de factores de transcripción no puedan interactuar y no puedan inducir la transcripción del péptido inductor de la lisis celular. En una célula enferma, en donde se ha producido un evento de translocación, el gen Bcr y el gen Abl se translocan de manera que los genes estén muy juntos. En este caso, las secuencias de ácido nucleico diana están lo suficientemente juntas como para que las dos partes inactivas del sistema del factor de transcripción se unan para inducir la transcripción del péptido inductor de la muerte celular. La lisis celular dependerá de un genotipo particular resultante del evento de movilidad genética.

EJEMPLO 18: Uso de un polipéptido dirigido a un sitio como agente terapéutico para enfermedades que comprenden un evento de mutación genética

El sistema descrito en el Ejemplo 17 también puede usarse para detectar la presencia de una mutación específica dentro de una célula. En este ejemplo, se elige el primer ácido nucleico dirigido a ácido nucleico para dirigir el polinucleótido dirigido a un sitio a una secuencia nativa ubicada cerca de un sitio de mutación. El segundo ácido nucleico dirigido a ácido nucleico se elige para reconocer una secuencia mutante (por ejemplo, la secuencia mutante que se ha identificado mediante secuenciación de ADN). El ácido nucleico dirigido a ácido nucleico se elige de manera que la secuencia mutante se produzca dentro de los 12 primeros ácidos nucleicos inmediatamente 5' de la secuencia PAM en el sitio. En este caso, las secuencias de ácido nucleico diana están lo suficientemente juntas como para que las dos partes inactivas del factor de transcripción se unan para permitir la transcripción de un péptido inductor de la lisis celular.

EJEMPLO 19: Reclutamiento del sistema inmunitario para atacar el tejido enfermo que contiene un evento de movilidad genética o una mutación genética.

El sistema descrito en el Ejemplo 17 y/o 18 también se puede usar para dirigir la transcripción mediante el sistema de

factor de transcripción dividido que dará como resultado la visualización de un antígeno en la superficie celular. En algunos casos, el antígeno es un péptido presentado por moléculas de MHC de clase II. En algunos casos, el antígeno es una proteína de la superficie celular que recluta células efectoras inmunitarias en un sitio.

EJEMPLO 20: Detección de la posición tridimensional de los ácidos nucleicos

Uno o varios vectores que comprenden un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y/o una proteína efectora se introduce en una célula. Una vez dentro de la célula, se forma un complejo que comprende los elementos codificados en el/los vector/es. Se introducen dos vectores en la célula. Un vector codifica una proteína efectora (Csy4) que comprende una primera parte inactiva de un sistema de marcador de afinidad dividido. Un segundo vector codifica una proteína efectora (Csy4, Cas5 o Cas6) que comprende una segunda parte inactiva del marcador de afinidad dividido. El ácido nucleico dirigido a ácido nucleico de los complejos se modifica con una secuencia de unión a proteínas Csy4, Cas5 o Cas6. Las proteínas efectoras se unen al ácido nucleico dirigido a ácido nucleico modificado. Los ácidos nucleicos dirigidos a ácido nucleico están diseñados para guiar los complejos a regiones de interés en una estructura de ácido nucleico tridimensional (por ejemplo, cromatina). Si las secuencias diana no están juntas en el espacio, las dos partes inactivas del marcador de afinidad dividido no pueden interactuar. Si las secuencias diana están juntas en el espacio, entonces las dos partes inactivas del marcador de afinidad dividido pueden unirse para formar el marcador de afinidad completo.

Las células se lisan y la lisis celular se incuba con un anticuerpo que se une al marcador de afinidad. El anticuerpo se purifica, purificando así el marcador de afinidad y el ácido nucleico al que se unen los complejos. El ácido nucleico purificado se disocia de los complejos usando un lavado con alto contenido de sal. El ácido nucleico purificado disociado se prepara para el análisis de secuenciación y se secuencia. Los resultados de la secuencia corresponden a regiones de cromatina que están juntas en un espacio tridimensional. Los resultados de la secuencia se pueden usar para comprender mejor la expresión génica y tratar la enfermedad.

EJEMPLO 21: INGENIERÍA MULTIPLEX DE GENOMA

Se introduce un vector que comprende un agente de dirección genético multiplexado que comprende módulos de ácido nucleico que comprenden un ácido nucleico dirigido a ácido nucleico y una secuencia de unión a endorribonucleasa en una célula. En algunos aspectos, la célula ya comprende un polipéptido dirigido a un sitio y una endorribonucleasa. En algunos casos, la célula se pone en contacto con un vector que comprende una secuencia de polinucleótido que codifica un polipéptido dirigido a un sitio y un vector que comprende una secuencia de polinucleótido que codifica una endorribonucleasa. En algunos casos, la célula se pone en contacto con un vector que comprende una secuencia de polinucleótido que codifica tanto el polipéptido dirigido a un sitio como la endorribonucleasa. En algunos aspectos, el vector comprende una secuencia de polinucleótido que codifica una o más endorribonucleasas. En algunos aspectos, el vector comprende una secuencia polinucleotídica que codifica un agente de dirección genético multiplexado, un polipéptido dirigido a un sitio y una o más endorribonucleasas. La matriz se transcribe en ARN. La una o más endorribonucleasas se unen a una o más secuencias de unión a endorribonucleasas en el agente de dirección genético multiplexado. La una o más endorribonucleasas escinden la una o más secuencias de unión a endorribonucleasas en el agente de dirección genético multiplexado, liberando así los módulos individuales de ácido nucleico. En algunos aspectos, los módulos de ácido nucleico comprenden la totalidad, una parte o nada de la secuencia de unión a endorribonucleasa.

Los módulos de ácido nucleico liberados se unen a polipéptidos dirigidos a un sitio, formando así complejos. Los complejos están dirigidos a uno o más ácidos nucleicos diana. El uno o más módulos de ácido nucleico se hibridan con uno o más ácidos nucleicos diana. El uno o más polipéptidos dirigidos a un sitio escinden el uno o más ácidos nucleicos diana en un sitio de escisión definido por el módulo de ácido nucleico, dando como resultado uno o más ácidos nucleicos diana modificados.

En algunos aspectos, uno o más polinucleótidos y/o vectores donantes que codifican los mismos se introducen en la célula. Se incorporan uno o más polinucleótidos donantes en el uno o más ácidos nucleicos diana escindidos, dando lugar a uno o más ácidos nucleicos diana modificados (por ejemplo, adición). En algunos casos, el mismo polinucleótido donante se incorpora a múltiples sitios de escisión. En algunos casos, se incorporan uno o más polinucleótidos donantes a múltiples sitios de escisión. En algunos casos, no se introduce en las células ningún polinucleótido ni/o vector donante que lo codifique. En estos casos, el ácido nucleico diana modificado puede comprender una eliminación.

EJEMPLO 22: MÉTODO DE ADMINISTRACIÓN ESTOQUIOMÉTRICA DE ARN A UNA CÉLULA

En algunos aspectos, la divulgación proporciona un método para la administración estequiométrica de ácidos nucleicos al núcleo de una célula. En algunas realizaciones, se usan tres ácidos nucleicos administrables estequiométricamente: uno que codifica Cas9, uno que codifica un ácido nucleico dirigido a ácido nucleico y otro que codifica Csy4. Cada uno de los tres ácidos nucleicos comprende un sitio de unión a Csy4.

En algunos aspectos, el método proporciona un polipéptido de fusión en tándem. El polipéptido de fusión comprende

tres polipéptidos Csy4. Los tres polipéptidos Csy4 están separados por un enlazador. Los tres polipéptidos Csy4 se unen a los sitios de unión a Csy4 en cada una de las tres moléculas de ácido nucleico, formando así un complejo.

En algunos aspectos, el complejo se forma fuera de una célula y se introduce en la célula. El complejo se forma mezclando los tres ácidos nucleicos administrables estequiométricamente y la proteína de fusión, y dejando que ocurra la reacción para permitir la unión entre el polipéptido de fusión en tándem y los tres sitios de unión a Csy4. El complejo se introduce por inyección, electroporación, transfección, transformación, transducción vírica y similares. En el interior de la célula, se traducen algunos de los ácidos nucleicos del complejo. En algunos aspectos, los productos de traducción resultantes son Csy4 y NLS-Cas9 (por ejemplo, Cas9 que comprende una NLS. La NLS puede que no tenga que estar en el extremo N). El Csy4 escinde el sitio de unión a Csy4 en el ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico, liberando así el ácido nucleico dirigido a ácido nucleico del polipéptido de fusión en tándem. NLS-Cas9 se une al ácido nucleico dirigido a ácido nucleico liberado, formando así una unidad. Esta unidad se transloca al núcleo. Dentro del núcleo, la unidad es guiada a un ácido nucleico diana que se hibrida con el espaciador del ácido nucleico dirigido a ácido nucleico. La Cas9 de la unidad escinde el ácido nucleico diana. La escisión del ácido nucleico diana por Cas9 se denomina ingeniería genómica.

En algunos aspectos, el complejo se forma dentro de una célula. Se introduce en la célula un vector que codifica los tres ácidos nucleicos administrables estequiométricamente. Se introducen tres vectores diferentes que codifican uno de cada uno de los tres ácidos nucleicos administrables estequiométricamente en la célula. Se introducen dos vectores en la célula, en donde uno de los dos vectores codifica dos ácidos nucleicos administrables estequiométricamente y uno de los dos vectores codifica un ácido nucleico administrable estequiométricamente. Cualquiera de los vectores puede codificar el polipéptido de fusión en tándem.

En el interior de la célula, el ácido nucleico del vector que codifica el ARN o un polipéptido se transcribe en ARN. Se forma un complejo que comprende los tres ácidos nucleicos y el polipéptido de fusión en tándem, mediante el que cada una de las proteínas de unión a Csy4 se une a un sitio de unión a Csy4 en cada uno de los tres ácidos nucleicos. Se traducen los ácidos nucleicos del complejo. En algunos aspectos, los productos de traducción resultantes son Csy4 y NLS-Cas9 (por ejemplo, Cas9 que comprende una NLS. La NLS puede que no tenga que estar en el extremo N). El Csy4 escinde el sitio de unión a Csy4 en el ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico, liberando así el ácido nucleico dirigido a ácido nucleico del polipéptido de fusión en tándem. NLS-Cas9 se une al ácido nucleico dirigido a ácido nucleico liberado, formando así una unidad. Esta unidad se transloca al núcleo. Dentro del núcleo, la unidad es guiada a un ácido nucleico diana que se hibrida con el espaciador del ácido nucleico dirigido a ácido nucleico. La Cas9 de la unidad escinde el ácido nucleico diana.

EJEMPLO 23: MÉTODO DE ADMINISTRACIÓN ESTOQUIOMÉTRICA DE MÚLTIPLES ÁCIDOS NUCLEICOS DIRIGIDOS A ÁCIDO NUCLEICO A UNA CÉLULA

En algunos aspectos, la divulgación proporciona un método para la administración estequiométrica de una pluralidad de ácidos nucleicos a una célula, en donde parte de la pluralidad de ácidos nucleicos son ácidos nucleicos dirigidos a ácido nucleico. En algunos aspectos, la pluralidad de ácidos nucleicos comprende cuatro ácidos nucleicos: uno que codifica Cas9, dos que codifican ácidos nucleicos dirigidos a ácido nucleico y uno que codifica Csy4.

Los ácidos nucleicos pueden comprender dos o más sitios de unión de proteínas de unión a ácido nucleico. En algunos casos, el primer sitio de unión de proteínas de unión a ácido nucleico (por ejemplo, el sitio más 5') comprende un sitio de unión a Csy4. En algunos casos, el segundo sitio de unión de proteínas de unión a ácido nucleico (por ejemplo, el sitio más 3') comprende un sitio de unión de proteínas de unión a ácido nucleico diferente (por ejemplo, sitio de unión a MS2). En algunos casos, el segundo sitio de unión de proteínas de unión a ácido nucleico de cada uno de los ácidos nucleicos administrables estequiométricamente es diferente. Por ejemplo, el segundo sitio de unión de proteínas de unión a ácido nucleico puede ser un sitio que se une a uno de un polipéptido CRISPR (por ejemplo, Cas5, Cas6). En algunos casos, el ácido nucleico que codifica Cas9 también comprende una señal de localización nuclear (NLS).

En algunos aspectos, el polipéptido de fusión en tándem comprende tres proteínas de unión a ácido nucleico. Las tres proteínas de unión a ácido nucleico son Csy4, Cas5, Cas6. El polipéptido de fusión en tándem comprende una señal de localización nuclear. Los polipéptidos de fusión en tándem comprenden más de una copia de una proteína de unión a ácido nucleico (por ejemplo, 2 copias de Csy4, una copia de Cas5, una copia de Cas6).

En algunos aspectos, un complejo que comprende la proteína de fusión en tándem y los cuatro ácidos nucleicos se forma fuera de una célula. El complejo se forma mezclando los cuatro ácidos nucleicos y la proteína de fusión en tándem y dejando que la reacción ocurra para permitir la unión entre el polipéptido de fusión en tándem y los cuatro sitios de unión a la proteína de unión a ácido nucleico. El complejo se introduce en la célula. La introducción se produce por transformación, transfección, transducción vírica, microinyección o electroporación, o cualquier técnica capaz de introducir biomoléculas a través de una membrana celular. El complejo se forma dentro de la célula (por ejemplo, tras la introducción de vectores que comprenden secuencias de ácido nucleico que codifican los ácidos nucleicos y la proteína de fusión en tándem).

En el interior de la célula, los ácidos nucleicos que codifican Csy4 y Cas9 se traducen, resultando en Csy4 y NLS-

Cas9 (por ejemplo, Cas9 que comprende una NLS. La NLS puede que no tenga que estar en el extremo N). Csy4 escinde el sitio de unión a Csy4 en los ácidos nucleicos que codifican los ácidos nucleicos dirigidos a ácido nucleico, liberándolos así del polipéptido de fusión en tándem.

NLS-Cas9 se une a los ácidos nucleicos dirigidos a ácido nucleico liberados, formando así una pluralidad de unidades. Las unidades pueden translocarse al núcleo. Dentro del núcleo, las unidades son guiadas a un ácido nucleico diana que se hibrida con el espaciador del ácido nucleico dirigido a ácido nucleico. La Cas9 de la unidad escinde el ácido nucleico diana.

EJEMPLO 24: MÉTODO DE ADMINISTRACIÓN ESTOQUIOMÉTRICA DE ARN Y UN POLINUCLEÓTIDO DONANTE

La divulgación proporciona un método de administración estequiométrica de componentes de ARN que se puede usar en el diseño del genoma. El método también puede comprender la administración de un polinucleótido donante que puede insertarse en un sitio de diseño del genoma.

La divulgación proporciona un método para la administración estequiométrica de una pluralidad de ARN y un polinucleótido donante a una célula. En algunos aspectos, la pluralidad de ARN comprende tres ARN y un ADN. En algunos casos, los tres ARN son: uno que codifica Cas9, uno que codifica Csy4 y uno que codifica un ácido nucleico dirigido a ácido nucleico. El ADN es un ADN que codifica un polinucleótido donante.

En algunos casos, los ARN comprenden una pluralidad de sitios de unión a proteínas de unión a ácidos nucleicos (por ejemplo, dos sitios de unión a proteínas de unión a ácidos nucleicos). El primer sitio de unión de proteínas de unión a ácido nucleico (por ejemplo, el sitio más 5') comprende un sitio de unión a Csy4. El segundo sitio de unión de proteínas de unión a ácido nucleico (por ejemplo, el sitio más 3' comprende un sitio de unión de proteínas de unión a ácido nucleico diferente). El segundo sitio de unión de proteínas de unión a ácido nucleico en cada uno de los ácidos nucleicos de la divulgación es diferente. Por ejemplo, el segundo sitio de unión a la proteína de unión a ácido nucleico se une a un polipéptido CRISPR (por ejemplo, Cas5, Cas6) y/o una proteína de unión a ADN (por ejemplo, una proteína de dedo de cinc). Los ácidos nucleicos del método también comprenden una secuencia que codifica una señal de localización nuclear (por ejemplo, el ARN que codifica Cas9 y el ADN que codifica un polinucleótido donante).

En algunos aspectos, el polipéptido de fusión en tándem comprende cuatro proteínas de unión a ácido nucleico (por ejemplo, proteínas de unión a ARN y proteínas de unión a ADN). En algunos casos, tres proteínas de unión a ácido nucleico son Csy4, Cas5, Cas6, y la cuarta proteína de unión a ácido nucleico es una proteína de unión a ADN (por ejemplo, proteína de dedo de cinc). En algunos casos, el polipéptido de fusión en tándem comprende una señal de localización nuclear.

En algunos aspectos, un complejo que comprende la proteína de fusión en tándem y los ácidos nucleicos (por ejemplo, tres ARN y un ADN) se forma fuera de una célula. El complejo se forma mezclando los ácidos nucleicos (por ejemplo, tres ARN y un ADN) y la proteína de fusión en tándem, y permitiendo que se produzca la reacción para permitir la unión entre el polipéptido de fusión en tándem y cuatro sitios de unión de la proteína de unión a ARN. El complejo se puede introducir en la célula. El complejo se forma dentro de la célula (por ejemplo, tras la introducción de vectores que comprenden secuencias de ácido nucleico que codifican los ácidos nucleicos y la proteína de fusión en tándem).

En el interior de la célula, los ARN que codifican Csy4 y Cas9 se traducen, resultando en Csy4 y NLS-Cas9 (por ejemplo, Cas9 que comprende una NLS. La NLS puede que no tenga que estar en el extremo N como se escribe aquí). Csy4 puede escindir el sitio de unión a Csy4 en los ARN que codifican el ácido nucleico dirigido a ácido nucleico y el ADN, liberándolos así del polipéptido de fusión en tándem. En algunos casos, el polinucleótido donante liberado es traslocalizado al núcleo por su señal de localización nuclear.

NLS-Cas9 se une al ácido nucleico dirigido a ácido nucleico liberado, formando así una unidad. La unidad puede translocarse al núcleo. Dentro del núcleo, la unidad es guiada a un ácido nucleico diana que se hibrida con el espaciador del ácido nucleico dirigido a ácido nucleico de la unidad. La Cas9 de la unidad escinde el ácido nucleico diana. El polinucleótido donante, una parte de un polinucleótido donante, una copia del polinucleótido donante o una parte de una copia del polinucleótido donante se puede insertar en el ácido nucleico diana escindido.

EJEMPLO 25: Selección perfecta de células modificadas genéticamente

Se pone en contacto una pluralidad de células con un vector que comprende secuencias que codifican un polipéptido homólogo a Cas9, un ácido nucleico dirigido a ácido nucleico y un polinucleótido donante. En algunos casos, una o más de las secuencias que codifican el polipéptido homólogo a Cas9, el ácido nucleico dirigido a ácido nucleico y el polinucleótido donante se encuentran en diferentes vectores, y las células se transfectan con el vector. En algunos casos, las células están infectadas con un virus que porta el vector. En algunos casos, la célula ya contiene una proteína homóloga a Cas9 y el vector no codifica este polipéptido. En algunos casos, la célula ya comprende un sistema CRISPR (por ejemplo, proteínas Cas, ARNcr y ARNcrtra) y el vector solo codifica el polinucleótido donante. El polinucleótido donante comprende secuencias que codifican un elemento genético de interés y un elemento indicador. El elemento indicador comprende secuencias de ácido nucleico dirigidas a ácido nucleico, una proteína

homóloga a Cas9 y una proteína fluorescente. El ácido nucleico dirigido a ácido nucleico guía a Cas9 a un ácido nucleico diana (por ejemplo, un sitio en el genoma de la célula hospedadora), lo que produce una rotura de ADN bicatenaria del ácido nucleico diana y la inserción del polinucleótido donante. La inserción del polinucleótido del donante se criba mediante la detección del indicador. En algunos casos, el cribado comprende la clasificación de células activadas por fluorescencia. El cribado comprende múltiples métodos de selección. Cas9 y/o los ácidos nucleicos dirigidos a ácido nucleico están controlados por un promotor inducible. Tras seleccionar una población de células que comprende la señal indicadora, el elemento indicador se elimina activando el promotor inducible, que transcribe los ácidos nucleicos dirigidos a ácido nucleico y el polipéptido dirigido a un sitio del polinucleótido donante. El ácido nucleico dirigido a ácido nucleico transcrito y el polipéptido dirigido a un sitio transcrito pueden formar complejos. Un complejo puede dirigirse al extremo 3' del elemento indicador del polinucleótido donante. Un complejo puede dirigirse al extremo 5' del elemento indicador del polinucleótido donante. Los extremos 3' y 5' del elemento indicador pueden escindirse. El ácido nucleico diana escindido puede volverse a unir por mecanismos celulares, dando como resultado una secuencia de ácido nucleico en fase que codifica la misma secuencia de ácido nucleico que antes de la inserción del polinucleótido donante. De esta manera, el elemento indicador se inserta perfectamente y se elimina de las células.

EJEMPLO 26: Método para la escisión de Cas9 usando ácidos nucleicos dirigidos a ácido nucleico diseñados

Reactivos

Se digirió el plásmido pCB002 que contiene la secuencia de ADN diana temp3 con 1 U de AscI por 1 ug de ADN para linealizar el vector. La reacción se detuvo mediante la incubación de la mezcla de reacción a 80 ºC durante 20 min. La reacción se purificó luego usando el kit de limpieza Qiagen de PCR.

El ácido nucleico guía sencillo se generó usando el kit de síntesis de ARN de alto rendimiento de T7 (n.º de cat. E2040S), usando la mitad del volumen de reactivos recomendados por los fabricantes para ARN de > 300 nucleótidos. En general, se usaron entre 200-350 ng de plantilla en una reacción de 20 µl que se incuba durante 16 horas. Las muestras se trataron con ADNasa y se purificaron usando el kit de purificación de ARN Thermo GeneJet (n.º de cat. K0732), y se eluyeron en 20 ul. Los rendimientos típicos varían de 1,4 a 2 ug/ul.

Al comienzo de la configuración del ensayo de escisión, todos los ARNgs se diluyeron a una concentración de 3500 nM y se sometieron a choque térmico a 80 ºC durante 15 minutos. Las muestras se retiraron del elemento calefactor y se dejaron equilibrar a temperatura ambiente. Una parte alícuota de 4 µl de ácido nucleico dirigido a ácido nucleico guía sencillo puede extenderse en un gel de agarosa para confirmar la integridad del ARN.

Se retiraron alícuotas de Cas9 a 2-2,5 mg/ml del congelador y se descongelaron lo más rápido posible, y luego se diluyeron en tampón de escisión x1 hasta la concentración de reserva apropiada.

Ensayo de escisión

Se formaron alícuotas de una mezcla maestra de agua, tampón de escisión x5 (HEPES 100 mM, KCl 500 mM, $MgCl_2$ 25 mM, DTT 5 mM y glicerol al 25 % a pH 7,4), y Cas9 a 250 nM en tubos de PCR de pared fina. Se añadió ARNgs a los tubos apropiados a una concentración final de 250 nM. La reacción se incubó a 37 ºC durante 10 minutos, se añadió plásmido linealizado 10 nM y se incubaron las reacciones (volumen de reacción final de 20 ul) a 37 ºC durante y hora. La reacción se terminó calentando la reacción a 60 ºC durante 20 minutos. Se mezclaron alícuotas de 10 ul con 2 ul de colorante de carga de ADN x6, y se analizaron por electroforesis en un gel de agarosa al 1,5 % teñido con SYBR seguro. La aparición de fragmentos de ~2800 pb y ~1300 pb fue indicativa de la escisión mediada por Cas9.

Los resultados de los experimentos se muestran en la Figura 23A, Figura 23B y Figura 23C. Todas las secuencias de ARN guía sintéticas diseñadas y mostradas en la Figura 22A y la Figura 22B soportaron la escisión de ARNgs, excepto SGRv8, en donde se invirtió toda la región complementaria. Estos resultados indican que diferentes regiones del ARNgs pueden ser susceptibles de modificarse y aún conservan la función.

Los ácidos nucleicos dirigidos a ácido nucleico diseñados se probaron en el ensayo descrito anteriormente. Las **Figuras 22A y 22B** muestran los diseños de un conjunto inicial de variantes de dúplex en variantes de cadena principal de ácido nucleico dirigido a ácido nucleico guía sencillo usadas para probar la actividad de escisión y dirección.

La **Figura 24** y la **Figura 25** ilustran un segundo conjunto de variantes de dúplex con modificaciones más pequeñas en el dúplex. V28 comprende una inserción de 2 bases 3' en la región complementaria; V29 comprende una eliminación de 3 bases 3' en la región complementaria.

La **Figura 26A** y la **Figura 26B** ilustran variantes de crtra que comprenden mutaciones en la parte de ARNcrtra del ácido nucleico dirigido a ácido nucleico (es decir, secuencia mínima de ARNcrtra y secuencia de ARNcrtra de 3'). V38-V41 comprenden fusiones entre la región complementaria/dúplex y los extremos 3' de las secuencias de ácido nucleico crtra de 163K de *M. mobile* (v38), FMD-9 de *S. thermophilus* (V39), *C. jejuni* (V40) y *N. meningitides*.

La **Figura 27A** y la **Figura 27B** representan variantes que comprenden modificaciones para permitir la unión de Csy4 al ácido nucleico dirigido a ácido nucleico. Se derivan secuencias de horquilla adicionales de la repetición CRISPR en PA14 de *Pseudomonas aeruginosa*.

La **Figura 23A,** la **Figura 23b** y la **Figura 23C** muestran datos del ensayo de escisión *in vitro* que demuestran la actividad de las variantes de ácido nucleico dirigido a ácido nucleico en la escisión de Cas9. La variante SGRv8 no pudo soportar la escisión del ácido nucleico diana (**Figura 23B**, carril 9).

La **Figura 28** y la **Figura 29** muestran dos repeticiones independientes del ensayo de escisión de Cas9 que prueba las variantes representadas en las **Figuras 24, 25, 26A, 26B, 27A,** y **27B.**

Se prepararon ácidos nucleicos dirigidos a ácido nucleico diseñados adicionales como se enumeran en la Tabla 5, y se probaron en el mismo ensayo. Los resultados del ensayo se muestran en las **Figuras 37A**, **37B, 37C,** y 37D, y se enumeran en la columna de actividad de la Tabla 5.

Los resultados de estos experimentos indican la importancia de las regiones de protuberancia y dominio P para efectuar la escisión del ácido nucleico diana. La funcionalidad de las variantes 42-45 indica que la adición de una secuencia de unión a Csy4 a un ácido nucleico dirigido a ácido nucleico no interrumpe la escisión del ácido nucleico diana.

EJEMPLO 27: SISTEMAS DE ANÁLISIS DE SECUENCIACIÓN

La **FIGURA 30** representa un sistema que está configurado para implementar los métodos de la divulgación. El sistema puede incluir un servidor informático ("servidor") que está programado para implementar los métodos descritos en el presente documento. La **FIGURA 30** representa un sistema **3000** adaptado para permitir que un usuario detecte, analice y comunique resultados de secuenciación de, por ejemplo, ácidos nucleicos enriquecidos dirigidos a nucleasas, ácidos nucleicos diana secuenciados, datos relativos a los métodos de divulgación, diagnostique una enfermedad, genotipe un paciente, tome una decisión de tratamiento específica del paciente, o cualquier combinación de los mismos. El sistema **3000** incluye un servidor informático central **3001** que está programado para implementar métodos ilustrativos descritos en el presente documento. El servidor **3001** incluye una unidad central de procesamiento (CPU, también "procesador") **3005** que puede ser un procesador de un solo núcleo, un procesador de múltiples núcleos o una pluralidad de procesadores para el procesamiento en paralelo. El servidor **3001** también incluye una memoria **3010** (por ejemplo, memoria de acceso aleatorio, memoria de solo lectura, memoria flash); unidad de almacenamiento electrónico **3015** (por ejemplo, disco duro); interfaz de comunicaciones **3020** (por ejemplo, adaptador de red) para comunicarse con uno o más sistemas; y dispositivos periféricos **3025** que pueden incluir memoria caché, otra memoria, almacenamiento de datos y/o adaptadores de presentación electrónica. La memoria **3010**, la unidad de almacenamiento **3015**, la interfaz **3020** y los dispositivos periféricos **3025** están en comunicación con el procesador **3005** a través de un bus de comunicaciones (líneas continuas), tales como una placa base. La unidad de almacenamiento **3015** puede ser una unidad de almacenamiento de datos para almacenar datos. El servidor **3001** está acoplado operativamente a una red informática ("red") **3030** con la ayuda de la interfaz de comunicaciones **3020**. La red **3030** puede ser Internet, una intranet y/o una extranet, una intranet y/o extranet que está en comunicación con Internet, una red de telecomunicaciones o datos. La red **3030**, en algunos casos, con la ayuda del servidor **3001**, puede implementar una red de igual a igual, que puede permitir que los dispositivos acoplados al servidor **3001** se comporten como un cliente o un servidor. El microscopio y el micromanipulador pueden ser dispositivos periféricos **3025** o sistemas informáticos remotos **3040**.

La unidad de almacenamiento **3015** puede almacenar archivos, tales como los resultados de la secuenciación, sitios de unión a dianas, datos genéticos personalizados, genotipos, imágenes, análisis de datos de imágenes y/o resultados de secuenciación, o cualquier aspecto de los datos asociados con la divulgación.

El servidor puede comunicarse con uno o más sistemas informáticos remotos a través de la red **3030**. El uno o más sistemas informáticos remotos pueden ser, por ejemplo, ordenadores personales, ordenadores portátiles, tablets, teléfonos, teléfonos inteligentes o asistentes digitales personales.

En algunas situaciones, el sistema **3000** incluye un único servidor **3001**. En otras situaciones, el sistema incluye múltiples servidores en comunicación entre sí a través de una intranet, extranet y/o Internet.

El servidor **3001** se puede adaptar para almacenar resultados de secuenciación, sitios de unión a dianas, datos genéticos personalizados y/u otra información de posible relevancia. Dicha información puede almacenarse en la unidad de almacenamiento **3015** o en el servidor **3001** y dichos datos pueden transmitirse a través de una red.

Los métodos descritos en el presente documento pueden implementarse mediante un código ejecutable (o software) de la máquina (o procesador informático) almacenado en una ubicación de almacenamiento electrónico del servidor **3001**, tal como, por ejemplo, en la memoria **3010** o la unidad de almacenamiento electrónico **3015**. Durante su uso, el procesador **3005** puede ejecutar el código. En algunos casos, el código puede recuperarse de la unidad de almacenamiento **3015** y almacenarse en la memoria **3010** para que el procesador **3005** pueda acceder fácilmente. En

algunas situaciones, la unidad de almacenamiento electrónico **3015** puede excluirse, y las instrucciones ejecutables por la máquina se almacenan en la memoria **3010**. Como alternativa, el código puede ejecutarse en un segundo sistema informático **3040**.

Los aspectos de los sistemas y métodos proporcionados en el presente documento, tales como el servidor **3001**, pueden incorporarse en programación. Varios aspectos de la tecnología pueden considerarse "productos" o "artículos de fabricación" normalmente en forma de código ejecutable por la máquina (o procesador) y/o datos asociados que se llevan a cabo o se incorporan en un tipo de medio de lectura automática. El código ejecutable automáticamente puede almacenarse en una unidad de almacenamiento electrónico, tal como la memoria (por ejemplo, memoria de solo lectura, memoria de acceso aleatorio, memoria flash) o un disco duro. Los medios de tipo "almacenamiento" pueden incluir una parte o la totalidad de la memoria tangible de los ordenadores, procesadores o similares, o módulos asociados de los mismos, tales como varias memorias de semiconductores, unidades de cinta, unidades de disco y similares, que pueden proporcionar almacenamiento no transitorio en cualquier momento para la programación del software. Todo o partes del software a veces se pueden comunicar a través de Internet u otras redes de telecomunicaciones. Dichas comunicaciones, por ejemplo, pueden permitir la carga del software desde un ordenador o un procesador a otro, por ejemplo, desde un servidor de administración u ordenador Host a la plataforma informática de un servidor de aplicaciones. Así pues, otro tipo de medio que puede contener los elementos del software incluye ondas ópticas, eléctricas y electromagnéticas, tales como las que se usan a través de interfaces físicas entre dispositivos locales, a través de redes fijas alámbricas y ópticas, y a través de varios enlaces aéreos. Los elementos físicos que transportan dichas ondas, tales como los alámbricos o inalámbricos, los enlaces ópticos o similares, también pueden considerarse medios portadores del software. Como se usa en el presente documento, a menos que estén restringidos a medios de "almacenamiento" tangibles no transitorios, las expresiones tales como "medio legible" por ordenador o máquina se refieren a cualquier medio que participe en proporcionar instrucciones a un procesador para su ejecución.

Por lo tanto, un medio de lectura automática, tal como el código ejecutable por un ordenador, puede adoptar muchas formas, incluyendo, pero sin limitación, un medio de almacenamiento tangible, un medio de onda portadora o un medio de transmisión físico. Los medios de almacenamiento no volátiles pueden incluir, por ejemplo, discos ópticos o magnéticos, tal como cualquiera de los dispositivos de almacenamiento en cualquier ordenador o similares, que se pueden usar para implementar el sistema. Los medios de transmisión tangibles pueden incluir: cables coaxiales, cables de cobre y fibra óptica (incluyendo los cables que comprenden un bus dentro de un sistema informático). Los medios de transmisión de onda portadora pueden tomar la forma de señales eléctricas o electromagnéticas, u ondas acústicas o de luz, como las generadas durante las comunicaciones de datos por radiofrecuencia (RF) e infrarrojos (IR). Las formas comunes de medios legibles por ordenador incluyen, por tanto, por ejemplo: un disquete, un disco flexible, disco duro, cinta magnética, cualquier otro medio magnético, un CD-ROM, DVD, DVD-ROM, cualquier otro medio óptico, tarjetas perforadas, cinta de papel, cualquier otro medio de almacenamiento físico con patrones de orificios, una RAM, una ROM, una PROM y una EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, una onda portadora que transmita datos o instrucciones, cables o enlaces que transporten dicha onda portadora, o cualquier otro medio desde el cual un ordenador pueda leer código y/o datos de programación. Muchas de estas formas de medios legibles por ordenador pueden estar implicadas en llevar una o más secuencias de una o más instrucciones a un procesador para su ejecución.

EJEMPLO 28: SECUENCIACIÓN BASADA EN MATRICES USANDO UN POLIPÉPTIDO DIRIGIDO A UN SITIO.

Una muestra de ácido nucleico se liga con un marcador de ácido nucleico que comprende un solo ARN guía y un marcador detectable. En conjunto, la muestra de ácido nucleico ligada al marcador de ácido nucleico se denomina muestra de prueba marcada. La muestra de prueba marcada se pone en contacto con una micromatriz que comprende oligonucleótidos inmovilizados. Los oligonucleótidos inmovilizados son un banco de ácidos nucleicos bicatenarios. Los oligonucleótidos comprenden un marcador detectable (por ejemplo, marcador fluorescente). Los miembros individuales de la muestra de prueba marcada pueden hibridarse con los oligonucleótidos con los que comparten suficiente complementariedad para facilitar la hibridación. La cantidad de hibridación puede cuantificarse comparando las intensidades de los dos marcadores detectables del banco de muestras y los oligonucleótidos inmovilizados. Por ejemplo, los oligonucleótidos hibridados pueden mostrar dos marcadores detectables (el del banco de muestras y el oligonucleótido). Los oligonucleótidos no hibridados pueden mostrar un marcador detectable (el del oligonucleótido). Las sondas hibridadas se ponen en contacto con Cas9. Cas9 escinde los oligonucleótidos de la micromatriz que se han hibridado con miembros de la muestra de prueba marcada. La escisión por el polipéptido dirigido a un sitio permite que los miembros hibridados de la muestra de prueba marcada sean eliminados. Tras la escisión por el polipéptido dirigido a un sitio, solo quedan en la micromatriz los marcadores detectables de oligonucleótidos no hibridados. El marcador detectable restante se cuantifica. La cuantificación de los marcadores detectables restantes se correlaciona con qué secuencias se representaron en la muestra de ácido nucleico y cuáles no (por ejemplo, mediante cartografía de posición). Los oligonucleótidos que no muestran un marcador detectable restante corresponden a secuencias representadas en la muestra de ácido nucleico. Los oligonucleótidos que muestran un marcador detectable restante corresponden a secuencias que no estaban representadas en la muestra de ácido nucleico.

EJEMPLO 29: ESCISIÓN DE UN ÁCIDO NUCLEICO DIANA CON UN ÁCIDO NUCLEICO DIRIGIDO A ÁCIDO NUCLEICO MARCADO

Este ejemplo describe los resultados de la escisión del ácido nucleico diana con un ácido nucleico dirigido a ácido nucleico que comprende una secuencia de unión a Csy4 en el extremo 5' del ácido nucleico dirigido a ácido nucleico. Se incubó Cas9 con o sin un ARN guía dirigido a un solo sitio en una secuencia de ADN lineal bicatenaria. Después de 1 hora, se separaron los productos de escisión y se visualizaron en un gel de agarosa. La **Figura 13D** ilustra que la escisión de Cas9 mediada por un ácido nucleico dirigido a ácido nucleico marcado (carril 3) fue menos eficaz que la escisión de Cas9 mediada por un ácido nucleico dirigido a ácido nucleico sin marcar (carril 1). Después de 1 hora, aprox. el 100 % de la diana fue escindido por Cas9 guiada por el ácido nucleico dirigido a ácido nucleico sin marcar, donde solo fue escindida una pequeña fracción de la diana en el mismo tiempo por Cas9 guiada por el ácido nucleico dirigido a ácido nucleico marcado. Estos experimentos indican que se puede usar la ubicación de la secuencia no nativa para ajustar la eficacia de la escisión del complejo de Cas9: ácido nucleico dirigido a ácido nucleico. Por ejemplo, **La Figura 27A, Figura 27B,** y **Figura 29** muestran la funcionalidad de la adición de una secuencia de unión a Csy4 en varias ubicaciones del ácido nucleico dirigido a ácido nucleico que conservan la actividad.

EJEMPLO 30: GENES DE DISEÑO DEL GENOMA EN TRASTORNOS DE LA SANGRE

Un ácido nucleico dirigido a ácido nucleico que comprende una secuencia espaciadora descrita en la Tabla 2 se introduce en una célula con un polipéptido dirigido a un sitio, formando así un complejo. El complejo se dirige al gen que participa en un trastorno sanguíneo que tiene una complementariedad sustancial con la secuencia espaciadora del ácido nucleico dirigido a ácido nucleico. Una vez que el ácido nucleico dirigido a ácido nucleico se hibrida con el ácido nucleico diana, el polipéptido dirigido a un sitio escinde el ácido nucleico diana. El ácido nucleico diana escindido se puede diseñar con un polinucleótido donante.

EJEMPLO 31: PROTOCOLO PARA DETERMINAR LA ESCISIÓN Y LA MODIFICACIÓN DEL ÁCIDO NUCLEICO DIANA

Este protocolo puede usarse para determinar si un ácido nucleico diana se ha escindido o si el ácido nucleico diana se ha modificado, tal como con una inserción o eliminación. Los cebadores que rodean el sitio diana se usan para amplificar por PCR, en una reacción de 25 µl, un producto de 500-600 nt de ADNg. Los cebadores comprenden al menos 100 nt a cada lado del sitio de corte. Los productos resultantes del ensayo de escisión son de aproximadamente más de 100 nt.

Se procesan aproximadamente 5 µl de producto de PCR en un gel de agarosa para determinar si la amplificación está limpia. Con el producto de PCR restante, el protocolo de fusión e hibridación es el siguiente:

- 95 ºC 10 min
- de 95 °C a 85 °C (-2,0 °C/s)
- 85 °C 1 min
- de 85 °C a 75 °C (-0,3 °C/s)
- 75 °C 1 min
- de 75 °C a 65 °C (-0,3 °C/s)
- 65 °C 1 min
- de 65 °C a 55 °C (-0,3 °C/s)
- 55 °C 1 min
- de 55 °C a 45 °C (-0,3 °C/s)
- 45 °C 1 min
- de 45 °C a 35 °C (-0,3 °C/s)
- 35 ºC 1 min
- de 35 °C a 25 °C (-0,3 °C/s)
- 25 °C 1 min
- 4 °C Mantener

Se prepara una mezcla maestra T7E1 de agua, tampón NEB2 y enzima T7E1. Se multiplica por cada reacción, más extra. La Tabla 8 muestra los componentes de la mezcla maestra T7E1.

Tabla 8. Componentes de reacción para la mezcla maestra T7E1.

	Reacción x1
Agua	7,5 µl
Tampón NEB 2	2 µl
Enzima T7E1	0,5 µl
Producto de PCR	10 µl
Total	20 µl

A cada reacción, en un tubo de tapa de tira de 200 µl, se añaden los siguientes reactivos: mezcla maestra T7E1 (10 µl) y muestra de PCR (10 µl). La reacción se incuba a 37 ºC durante 25 minutos.

Se añade tampón de carga a la muestra y la muestra completa se procesa en un gel al 3 % a 120 V durante 20 minutos. El gel puede procesarse durante más tiempo si se requiere una mayor resolución. Se fotografía y se guarda la imagen del gel.

Se cuantifica la imagen para determinar la cantidad de escisión del ácido nucleico diana.

EJEMPLO 32: PRUEBA CELULAR DE VARIANTES DE ÁCIDO NUCLEICO DIRIGIDO A ÁCIDO NUCLEICO

Este ejemplo muestra que las variantes de ácido nucleico dirigido a ácido nucleico representadas en las ligaduras 22A, 22B, 24, 25 y el Ejemplo 26 se probaron en un ensayo basado en células para determinar si la funcionalidad *in vitro* determinada en el Ejemplo 26 coincidía con la funcionalidad *in vivo*. Las células HEK293 se cultivaron hasta 60-70 % de confluencia en placas de 10 cm. Las células se eliminaron por tripsinización, se contaron usando un hemocitómetro y luego se separaron en alícuotas de 7 x 10^4 células para cada pocillo por transfectar. Para cada pocillo, se mezclaron 250 ng de plásmido pCB045 que expresa Cas9 optimizada con codón de mamífero con 30 ng de ARN guía y 40 ng de ADN copGFP y 0,5 ul de Lipofectamina 2000 en un volumen final de 50 ul de DMEM. El ADN y los lípidos se incubaron durante 15 minutos antes de la transfección. Las transfecciones se llevaron a cabo en el momento de la colocación en placa, añadiendo la mezcla de lípido:ADN a 450 µl de DMEM + 10% de suero bovino fetal que contenía 7 x 10^4 células. La mezcla de transfección/células se añadió a placas de cultivo de tejidos de 96 pocillos recubiertas con colágeno de cola de rata I. Las células se incubaron a 37 ºC en una incubadora con CO2 al 5 % durante 40 horas.

Se retiraron los medios de cada uno de los pocillos y se lisaron las células usando solución Quickextract (Epicentro) de acuerdo con las instrucciones del fabricante. El ADN cosechado de la lisis con QuickExtract se diluyó 1:10 y se usó como plantilla en reacciones de PCR para ensayos de T7E1 como se describe en el Ejemplo 30.

La **Figura 36** indica que todas las variantes, excepto v8 y v9, pudieron escindir el ácido nucleico diana. La variante de ácido nucleico dirigida a ácido nucleico v8 también fue sustancialmente inactiva en ensayos *in vitro* tal como se representa en la **Figura 23B**. La variante de ácido nucleico dirigido a ácido nucleico v9 fue muy débilmente activa en el ensayo *in vitro* representado en la **Figura 23B**.

EJEMPLO 33: DETERMINACIÓN DEL DESTINO CELULAR CON UNA CÉLULA MARCADA

Este ejemplo describe cómo rastrear una célula que se desarrolla a partir de un linaje celular. Una célula madre hematopoyética (por ejemplo, Un hemocitoblasto) se pone en contacto con un polipéptido dirigido a un sitio, un ácido nucleico dirigido a ácido nucleico y un polinucleótido donante. El polipéptido dirigido a un sitio y el ácido nucleico dirigido a ácido nucleico forman un complejo y se dirigen a una región del genoma hematopoyético para la escisión. Una vez escindido, el polinucleótido donante se inserta en el sitio escindido Del genoma de la célula hepatopoyética. La célula madre hematopoyética se induce a diferenciarse mediante procesos de diferenciación normales. En diferentes etapas de diferenciación, se puede analizar la muestra que comprende las células hematopoyéticas diferenciadas para determinar la presencia del polinucleótido donante. De esta manera, se puede rastrear el proceso de diferenciación de una célula.

EJEMPLO 34: CLONACIÓN DE OLIGONUCLEOTIDO BICATENARIO QUE CODIFICA UN ÁCIDO NUCLEICO DIRIGIDO A ÁCIDO NUCLEICO EN UN VECTOR LINEALIZADO

El presente ejemplo describe cómo generar un oligonucleótido bicatenario que codifica una parte de ácido nucleico dirigido a ácido nucleico (por ejemplo, un espaciador) e insertarlo en un vector linealizado. El vector linealizado o un vector superenrollado cerrado comprende una secuencia que codifica un polipéptido dirigido a un sitio (por ejemplo, Cas9), un promotor que impulsa la expresión de la secuencia que codifica el polipéptido dirigido a un sitio (por ejemplo, promotor CMV), una secuencia que codifica un enlazador (por ejemplo, 2A), una secuencia que codifica un marcador (por ejemplo, CD4 u OFP), una secuencia que codifica la parte de un ácido nucleico dirigido a ácido nucleico, un promotor que impulsa la expresión de la secuencia que codifica una parte del ácido nucleico dirigido a ácido nucleico, y una secuencia que codifica un marcador seleccionable (por ejemplo, ampicilina), o cualquier combinación de las mismas.

Se recogen cantidades iguales de dos oligonucleótidos monocatenarios juntos (por ejemplo, 50 micromolar). Los dos oligonucleótidos monocatenarios pueden hibridarse entre sí. Al menos uno de los dos oligonucleótidos monocatenarios es complementario a un ácido nucleico diana (por ejemplo, una región de 10-30 nucleótidos adyacente a un motivo adyacente al protoespaciador en una diana). Al menos uno de los dos nucleótidos monocatenarios comprende una secuencia de saliente 3' que comprende la secuencia 5'-GTTT-3'. Al menos uno de los dos oligonucleótidos monocatenarios comprende un saliente 3’ que comprende la secuencia 5'-CGGTG-3'. En algunos casos, uno de los dos oligonucleótidos monocatenarios comprende un saliente 5'-GTTT-3’ y el otro de los dos oligonucleótidos monocatenarios comprende 5' -CGGTG-3’. El hibridación se realiza en un tampón de hibridación que comprende al menos Tris HCl 10 mM pH 8,0, EDTA1 mM, pH 8,0, y NaCl 100 mM. El hibridación se lleva a cabo calentando la mezcla de oligonucleótidos a 95 ºC durante 3-5 minutos, retirando la mezcla de oligonucleótidos de la fuente de calor, y dejando que la mezcla se enfríe a temperatura ambiente durante 5-10 minutos. La mezcla de oligonucleótidos bicatenarios se centrifuga suavemente. Tras la hibridación, la mezcla puede almacenarse a 4 ºC o -20 ºC. La mezcla, ahora de oligonucleótidos bicatenarios, se diluye para preparar dos soluciones madre de 500 nanomolar y 5 nanomolar. Las soluciones madre se preparan diluyendo la mezcla de oligonucleótidos en agua.

El oligonucleótido bicatenario (Oligonucleótidobc) se liga en un vector linealizado. El vector linealizado comprende una secuencia que codifica un polipéptido dirigido a un sitio (por ejemplo, Cas9), una proteína marcadora (por ejemplo, proteína fluorescente naranja) y/o una secuencia que codifica un ácido nucleico dirigido a ácido nucleico, en donde el vector linealizado está linealizado en una región de la secuencia que codifica el ácido nucleico dirigido a ácido nucleico, de modo que los extremos adhesivos generados coinciden con los extremos salientes del oligonucleótido bc. La reacción de ligación puede comprender un tampón de ligación x1 (por ejemplo, Tris-HCl 50 mM, pH 7,6, $MgCl_2$ 5 mM, ATP 1 mM, DTT 1 mM y/o PEG 8000 al 5 %), 30 nanogramos vector linealizado, Oligonucleótido bc 5 Nm y ADN ligasa (por ejemplo, 4 microlitros de tampón e ligadura x5, 2 microlitros de vector linealizado a 15 nanogramos/microlitro, 2 microlitros de Oligonucleótido bc 5 nanomolar, 11 microlitros de agua, 1 microlitro de ADN ligasa de T4). Se mezcla la reacción. La reacción se incuba a temperatura ambiente durante 2 horas y 10 minutos. La reacción se coloca en hielo y se transforma en células competentes.

La transformación en células competentes comprende la transformación en células TOP 10 *E. coli* químicamente competentes. Las células competentes se descongelan en hielo. Se añaden 3 microlitros de la mezcla de reacción a las células competentes y se mezclan suavemente. Las células se incuban en hielo durante 10-30 minutos. Las células son sometidas a choque térmico durante 30 segundos a 42 ºC. Las células se transfieren a hielo durante 2 minutos. Se añaden 250 microlitros de medio (SOC o LB) a las células. Las células se agitan a 200 rpm durante 1 hora a 37 ºC. Luego, las células se extienden sobre una placa de agar que comprende 100 microgramos/mililitro de ampicilina y se almacenan durante la noche a 37 ºC.

Se analizan los transformantes. Por ejemplo, los transformantes se analizan para determinar la identidad del Oligonucleótido bc ligado al vector y/o confirmar que la ligadura no es un falso positivo. Para analizar los transformantes, las colonias se recogen y se cultivan durante la noche en medio LB que comprende 100 microgramos/mililitro de amplicilina a 37 ºC. El plásmido que comprende el polipéptido dirigido a un sitio y el Oligonucleótido bc se aisla (por ejemplo, mediante un kit miniprep). Se realiza una reacción de secuenciación sobre el plásmido aislado. La reacción de secuenciación usa un cebador de secuenciación diseñado para secuenciar el Oligonucleótido bc (por ejemplo, el cebador de secuenciación es un cebador de secuenciación U6 que se une al promotor U6 que se encuentra cadena arriba de la secuencia que codifica el Oligonucleótido bc.

Una vez que se identifica la inserción deseada de un oligonucleótido bc, el plásmido se puede almacenar a -20 ºC o en una reserva de glicerol a -80 ºC. Para preparar una reserva de glicerol, se siembra en estrías la colonia original que comprende el plásmido deseado en una placa de agar que comprende 100 microgramos/mililitro de ampicilina y se incuba durante la noche a 37 ºC. Se aísla una sola colonia cultivada en LB que comprende 100 microgramos/mililitro de ampicilina hasta que el cultivo alcanza la fase estacionaria. El cultivo se mezcla con glicerol y se congela instantáneamente en nitrógeno líquido (por ejemplo, se mezclan 0,85 ml de cultivo con 0,15 ml de glicerol).

El plásmido purificado que comprende el Oligonucleótido bc deseado se inserta en una estirpe celular (por ejemplo, estirpe celular de mamífero, HeLa) mediante transfección. Para transfectar el plásmido, el plásmido se purifica a altas concentraciones usando, por ejemplo, un kit maxi prep. El plásmido se transfecta con tampón basado en lípidos (por ejemplo, Lipofectamina 2000) en las células que se siembran al 70 % de confluencia. Se transfectan 3 microgramos del vector en las células.

EJEMPLO 35: ADMINISTRACIÓN DE NANOPARTÍCULAS DE UN ÁCIDO NUCLEICO DIRIGIDO A ÁCIDO NUCLEICO DISEÑADO

Se preparará una nanopartícula que encapsula un ácido nucleico que codifica un ácido nucleico dirigido a ácido nucleico diseñado y un polipéptido dirigido a un sitio. Las nanopartículas se prepararán mezclando DOPE, Chol, DSPE-PEG y C_{16}mPEG-Ceramida en una proporción molar de 18:60:20:1:1 en 10 ml de etanol al 90 % (30 µmol de lípido total). El ácido nucleico se disolverá en 10 ml de tampón Tris 20 mM (pH 7,4-7,6). Tras calentar a 37 ºC, las dos soluciones se mezclarán entre sí a través de una bomba de jeringa de duelo, y la solución mezclada se diluirá

posteriormente con 20 ml de tampón Tris 20 mM (NaCl 300 mM, pH 7,4-7,6). La mezcla se incubará a 37 ºC durante 30 minutos y se dializará en tampón PBS 10 mM (NaCl 138 mM, KCl 2,7 mM, a pH 7,4). Se obtendrán partículas estables tras la eliminación del etanol de la mezcla por diálisis. La solución de nanopartículas se concentra por centrifugación a 3.000 rpm y una temperatura de 4 ºC.La suspensión concentrada se recogerá tras un tiempo determinado y se esterilizará por filtración a través de un filtro de jeringa de 0,22 µm (Millex-GV, Millipore, EE.UU.). Se obtendrá una suspensión homogénea de las nanopartículas que comprenden el ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico diseñado y el polipéptido dirigido a un sitio.

Se pondrán las nanopartículas en contacto con una célula. La nanopartícula entrará en la célula. En el interior de la célula, la nanopartícula liberará el ácido nucleico que codifica el ácido nucleico dirigido a ácido nucleico diseñado y el polipéptido dirigido a un sitio. El ácido nucleico se transcribirá y/o traducirá para producir un ácido nucleico dirigido a ácido nucleico diseñado que se une a una proteína de polipéptido dirigido a un sitio, formando así un complejo. El complejo se dirigirá a un ácido nucleico diana que se hibrida con el ácido nucleico dirigido a ácido nucleico diseñado. El complejo escindirá el ácido nucleico diana.

En algunos casos, la nanopartícula comprenderá además un ácido nucleico que codifica un polinucleótido donante. Cuando el ácido nucleico diana es escindido por el polipéptido dirigido a un sitio, el polinucleótido donante se insertará en el sitio del ácido nucleico diana escindido.

REIVINDICACIONES

1. Un ácido nucleico dirigido a ácido nucleico guía sencillo diseñado capaz de formar un complejo con una proteína Cas9 de CRISPR de Tipo II, comprendiendo el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado:

una secuencia espaciadora, que comprende una secuencia que puede hibridarse con la secuencia de ácido nucleico diana;
un dúplex que comprende

una región de hibridación entre una secuencia de repetición CRISPR mínima y una secuencia de crtra mínima, en donde la secuencia espaciadora está en 5' del dúplex y
una región de protuberancia que no está seguida por un primer tallo, en donde la región de protuberancia comprende una región de nucleótidos no emparejados, y en donde la región de protuberancia comprende una purina no emparejada en el lado de la secuencia de repetición CRISPR mínima de la región de protuberancia;

una secuencia de enlazador guía sencillo que conecta los extremos del dúplex para formar una estructura de horquilla; y
una secuencia de crtra de 3'.

2. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de la reivindicación 1, que comprende además una secuencia enlazadora que une el dúplex con un ARNcrtra medio, en donde el ARNcrtra medio comprende un segundo tallo y, opcionalmente, un tercer tallo.

3. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de la reivindicación 1 o 2, en donde la región de protuberancia está mutada.

4. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones 1 a 3, en donde la región de hibridación entre la secuencia de repetición mínima CRISPR y la secuencia de crtra mínima comprende no más de 2 emparejamientos erróneos.

5. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la secuencia de repetición CRISPR mínima tiene una longitud de entre 6 y 15 nucleótidos.

6. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la secuencia de repetición CRISPR mínima tiene una longitud de 12 nucleótidos.

7. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la secuencia de crtra mínima tiene una longitud de 14 nucleótidos.

8. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la región de protuberancia comprende, a un lado del dúplex, un 5'-XXXY-3' no emparejado, en donde X es cualquier purina e Y es un nucleótido que puede formar un par oscilante con un nucleótido en la cadena opuesta.

9. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la secuencia de protuberancia comprende cuatro nucleótidos no emparejados en un lado del dúplex.

10. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde la secuencia del enlazador guía sencillo tiene una longitud de 4 a 40 nucleótidos.

11. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de la reivindicación 10, en donde la secuencia del enlazador guía sencillo tiene una longitud de 3 a aproximadamente 10 nucleótidos.

12. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado comprende ADN.

13. El ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquiera de las reivindicaciones anteriores, en donde el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado comprende ARN.

14. Un polinucleótido que codifica el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de cualquier reivindicación anterior.

15. El polinucleótido de la reivindicación 14, que comprende además un promotor unido operativamente al polinucleótido.

16. Una composición que comprende:

el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de una cualquiera de las reivindicaciones 1 a 13; y
y una proteína Cas9,
en donde el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado y la proteína Cas9 forman un complejo.

17. La composición de la reivindicación 16, en donde la proteína Cas9 es una proteína Cas9 de *Streptococcus pyogenes*.

18. Un método *in vitro* para escindir un ácido nucleico diana, en donde el método no es un método de tratamiento del cuerpo humano o animal mediante cirugía o terapia realizadas en el cuerpo, en donde el método no es para modificar la identidad genética de la línea germinal del ser humano, que comprende:
poner en contacto un ácido nucleico que comprende el ácido nucleico diana con la composición de la reivindicación 16, facilitando así la unión del complejo al ácido nucleico diana, produciendo la escisión del ácido nucleico diana.

19. Un método *in vitro* de unión de un ácido nucleico diana, en donde el método no es un método de tratamiento del cuerpo humano o animal mediante cirugía o terapia realizadas en el cuerpo, en donde el método no es para modificar la identidad genética de la línea germinal del ser humano, que comprende:
poner en contacto un ácido nucleico que comprende el ácido nucleico diana con la composición de la reivindicación 16, facilitando así la unión del complejo al ácido nucleico diana.

20. Un kit que comprende:
el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado de una cualquiera de las reivindicaciones 1 a 13 o un polinucleótido que codifica el ácido nucleico dirigido a ácido nucleico guía sencillo diseñado; y un tampón.

21. El kit de la reivindicación 20, que comprende además:
una proteína Cas9 o un polinucleótido que codifica la proteína Cas9.

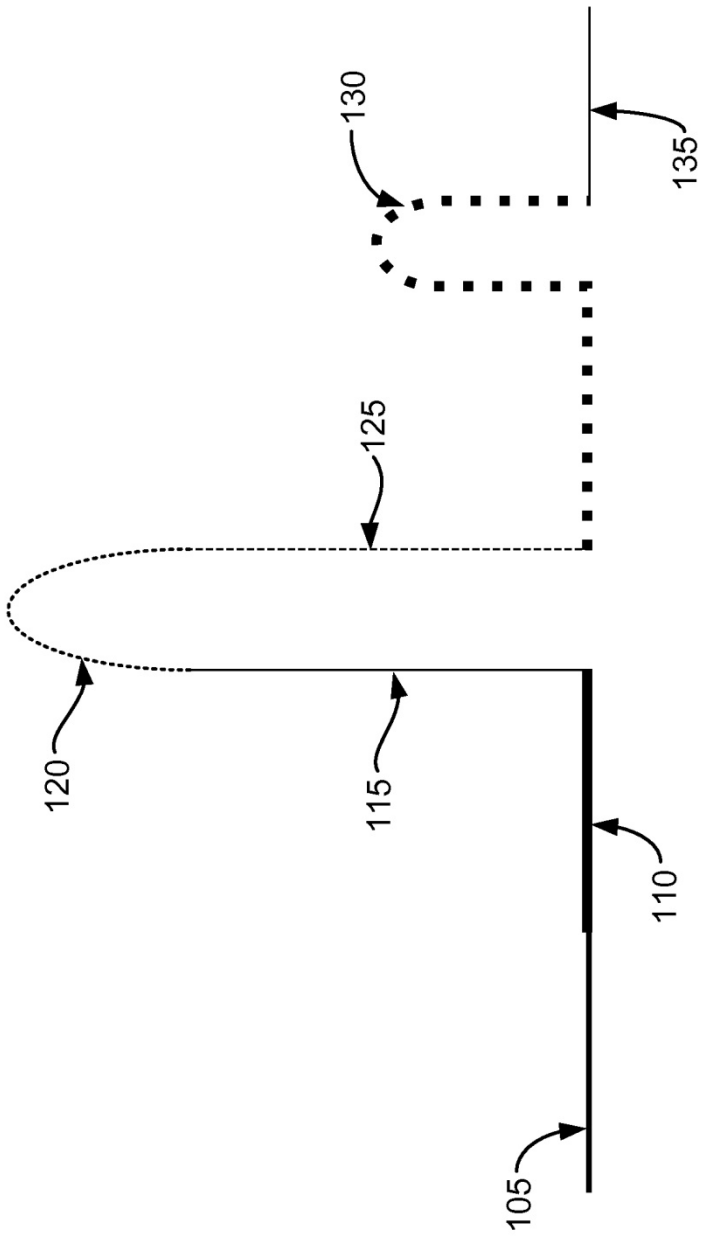


FIG. 1A

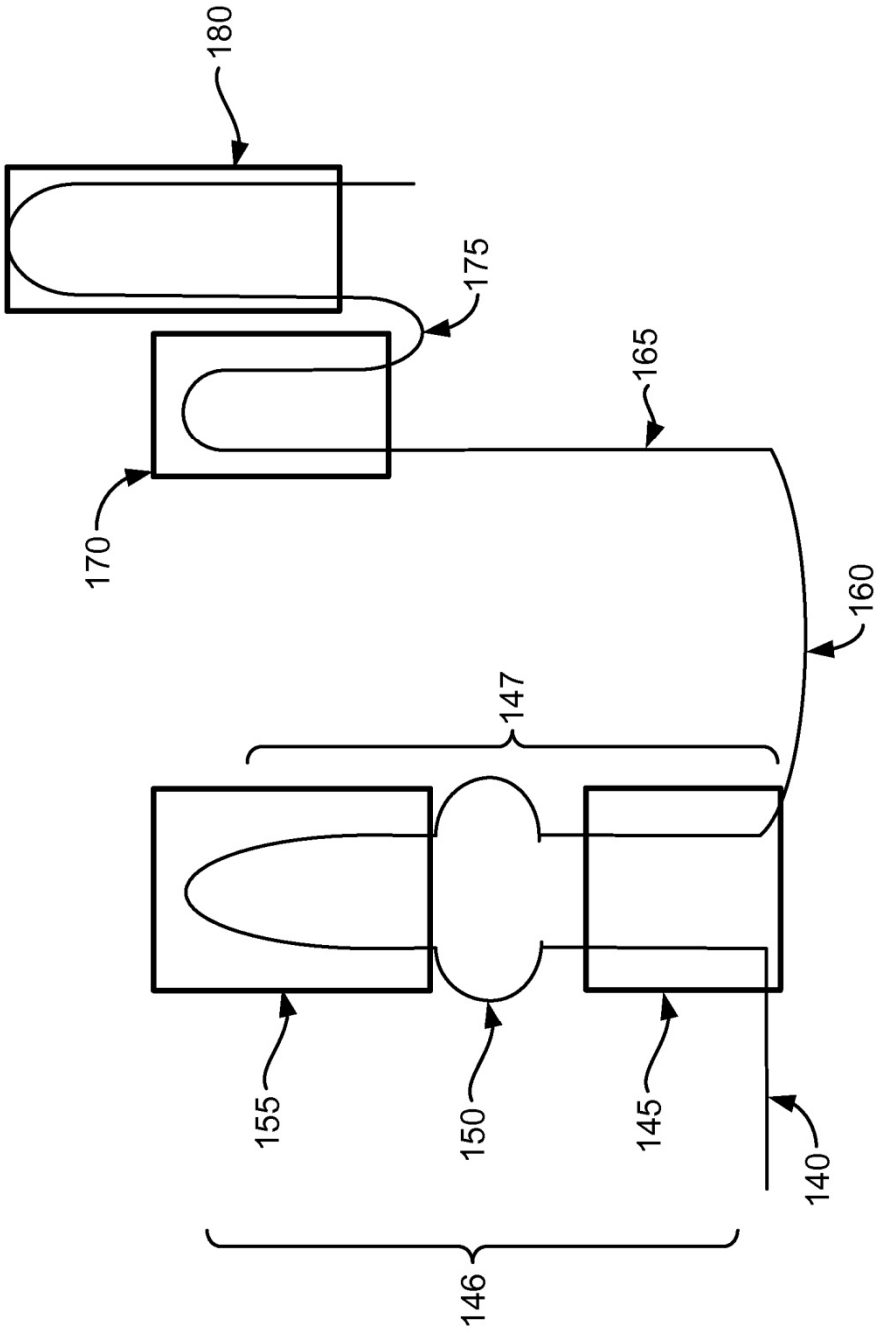


FIG. 1B

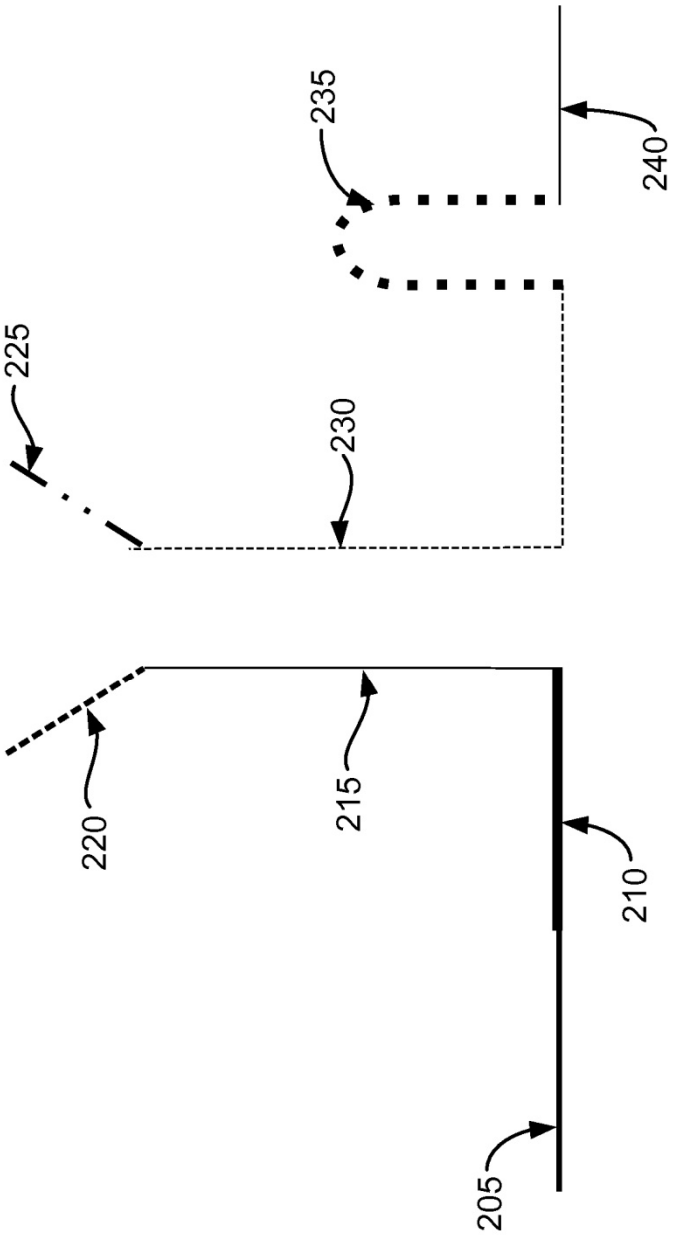


FIG. 2

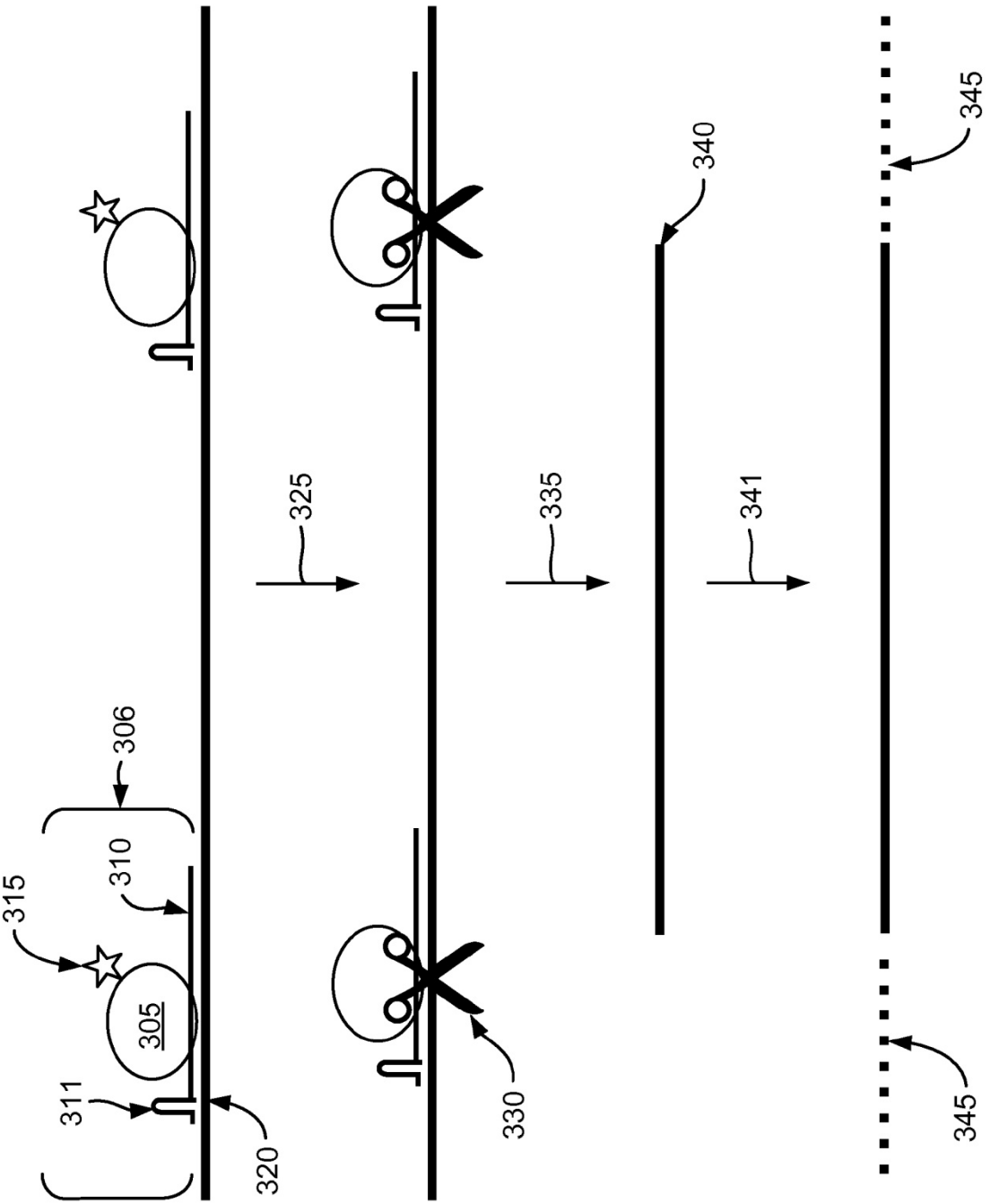


FIG. 3

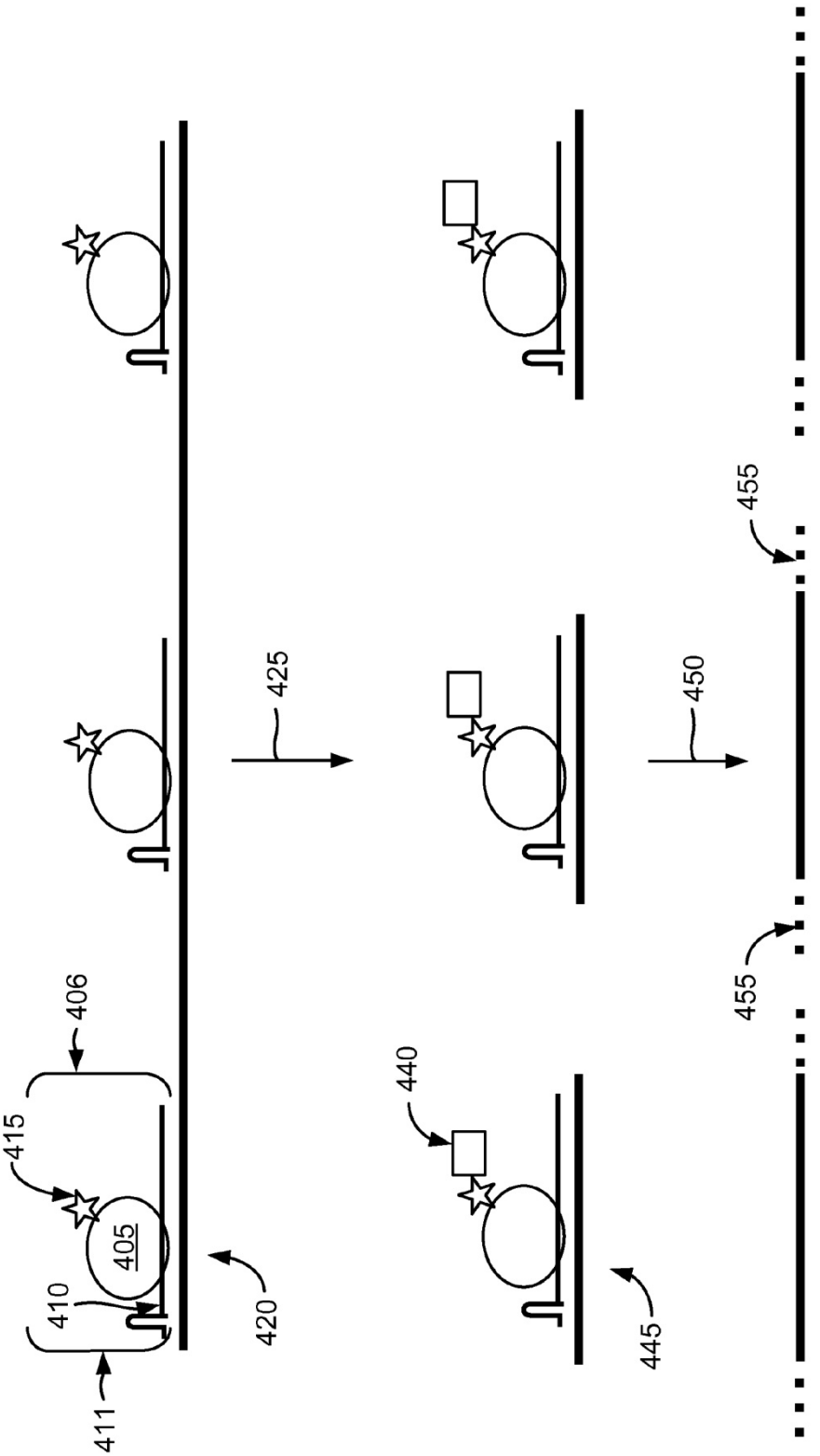


FIG. 4

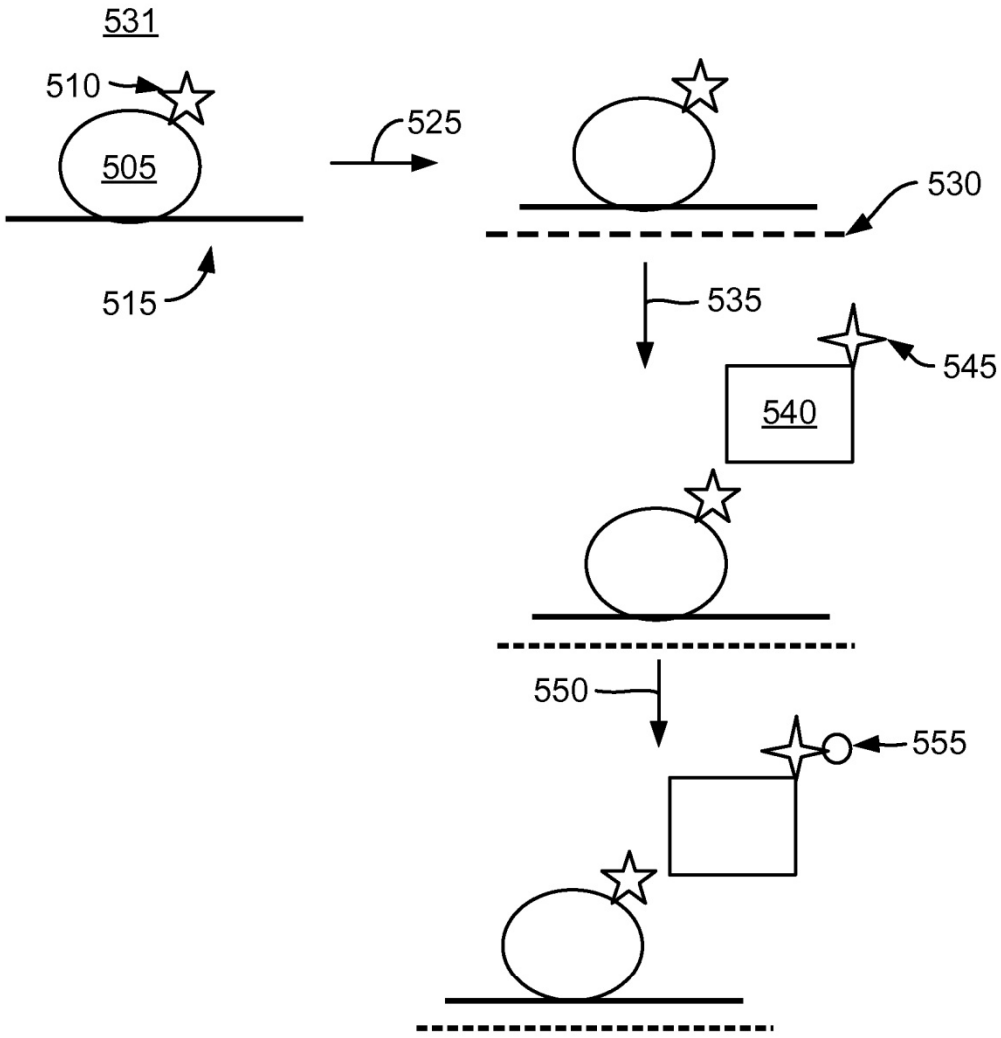


FIG. 5

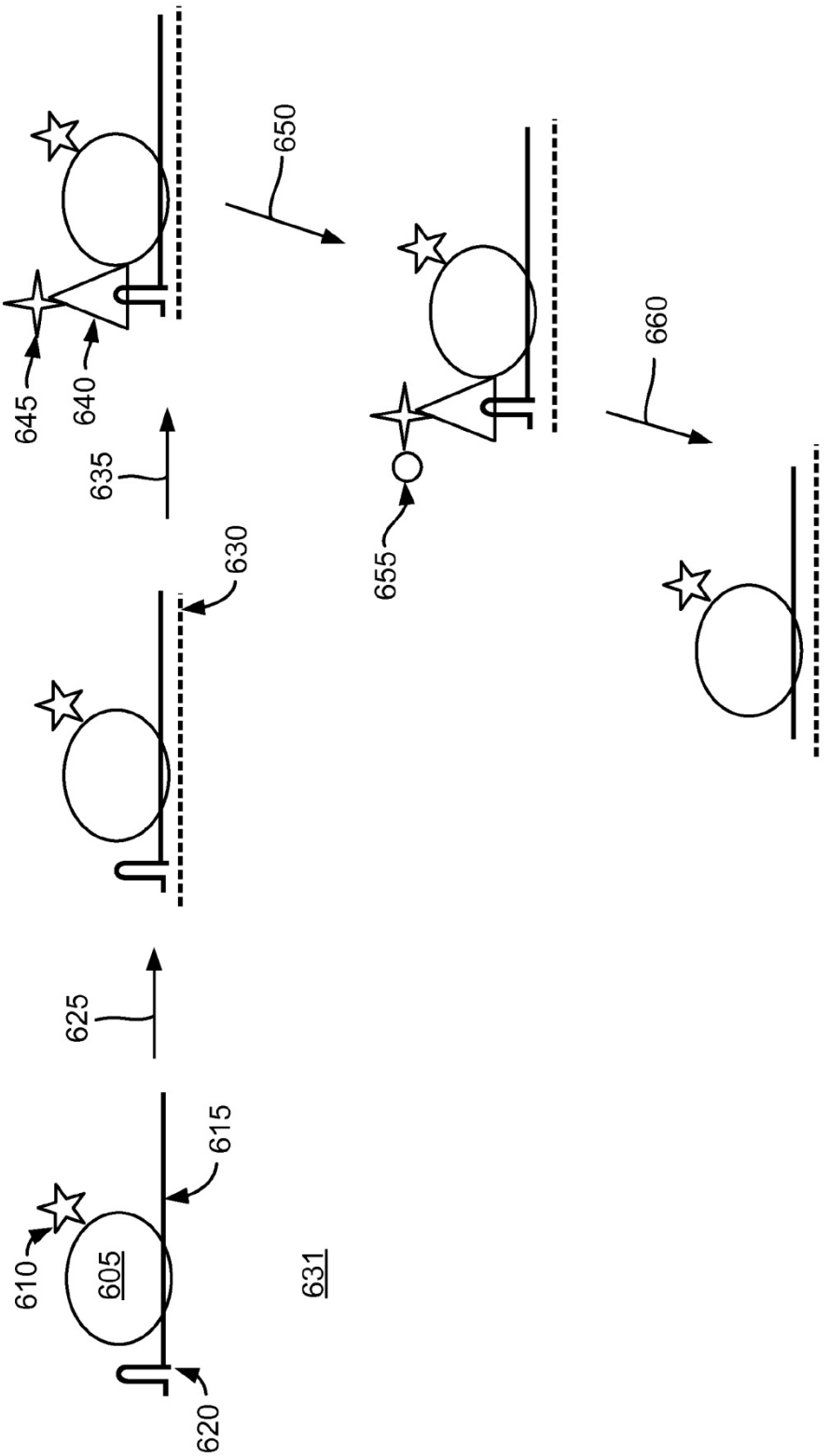


FIG. 6

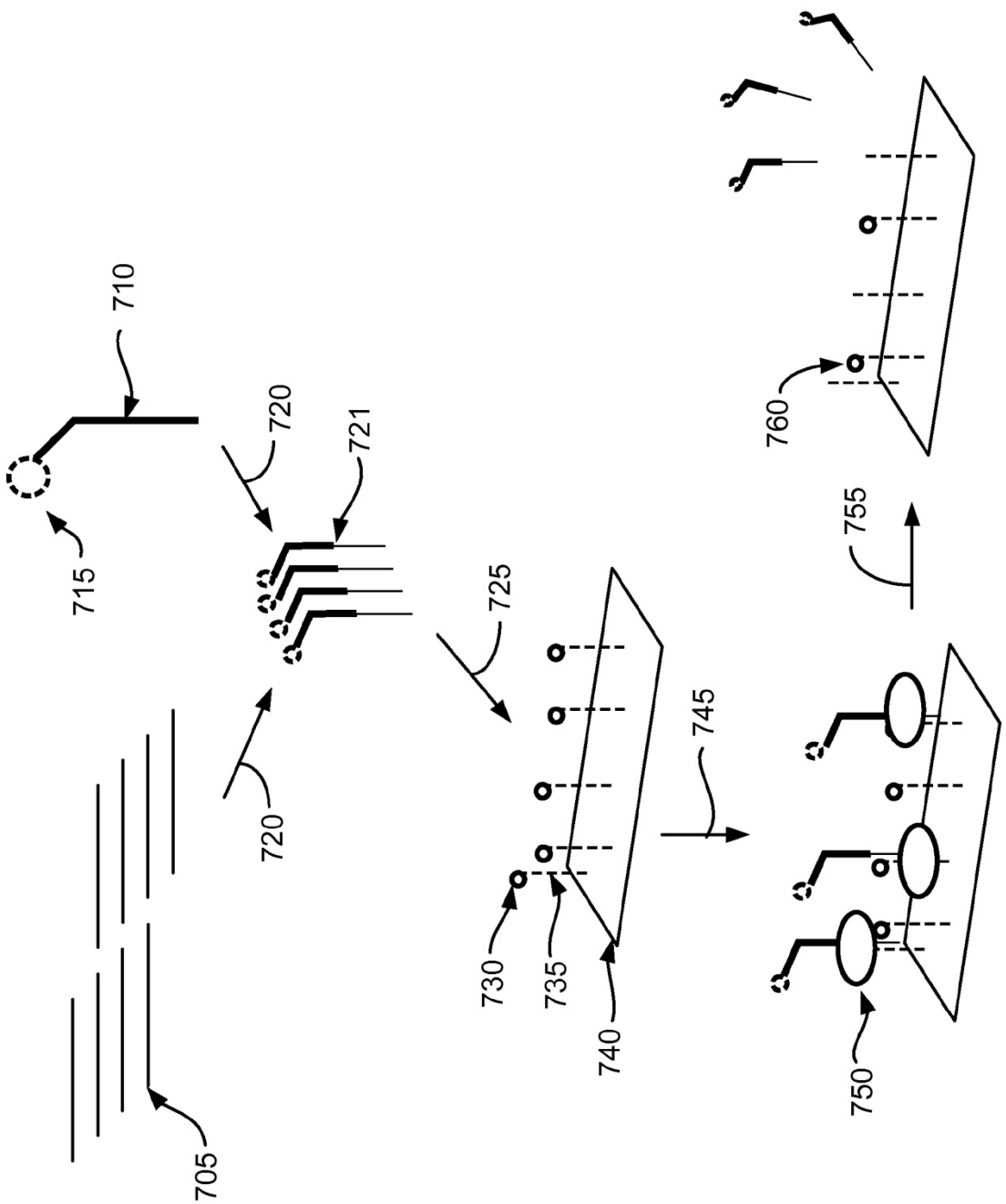


FIG. 7

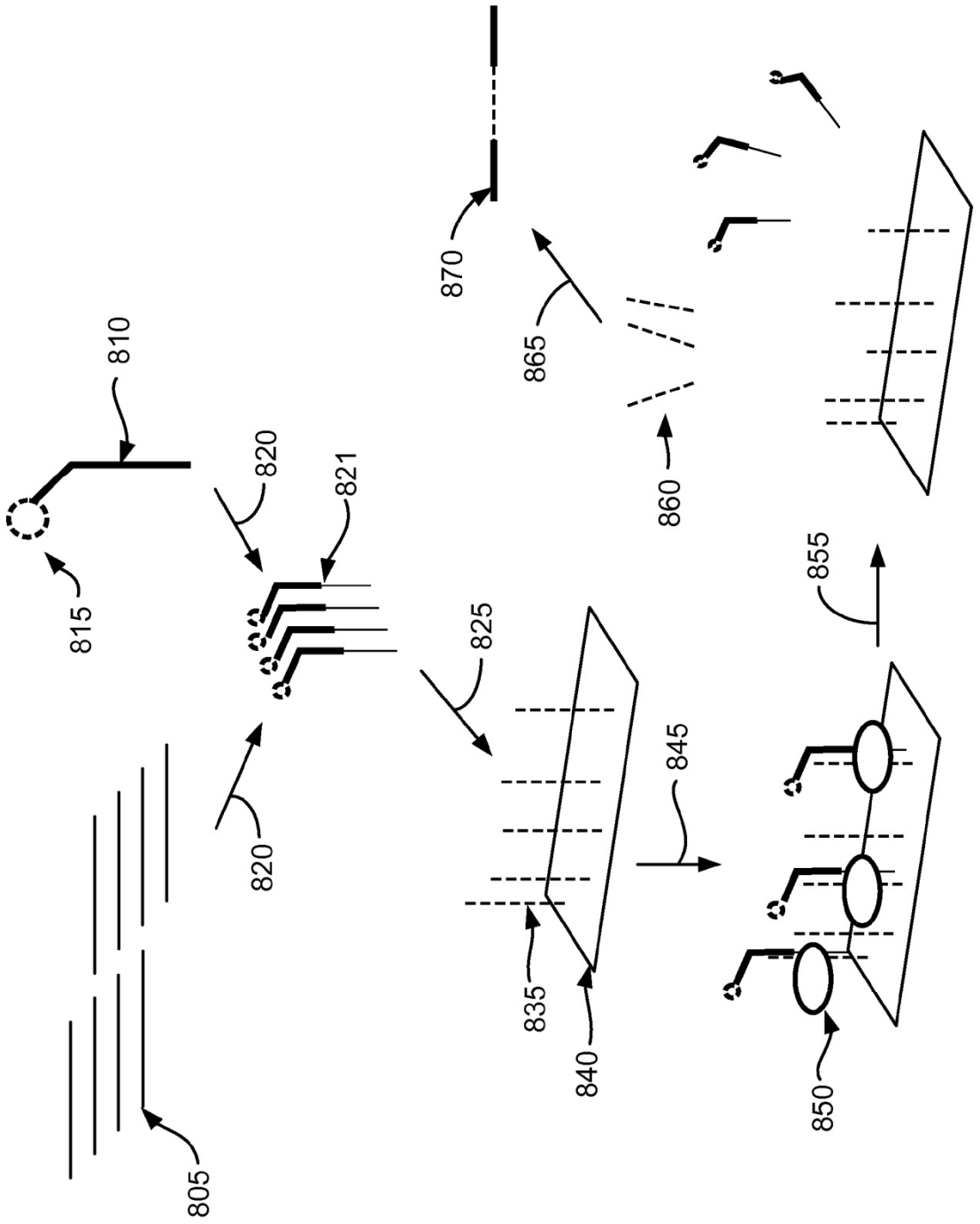


FIG. 8

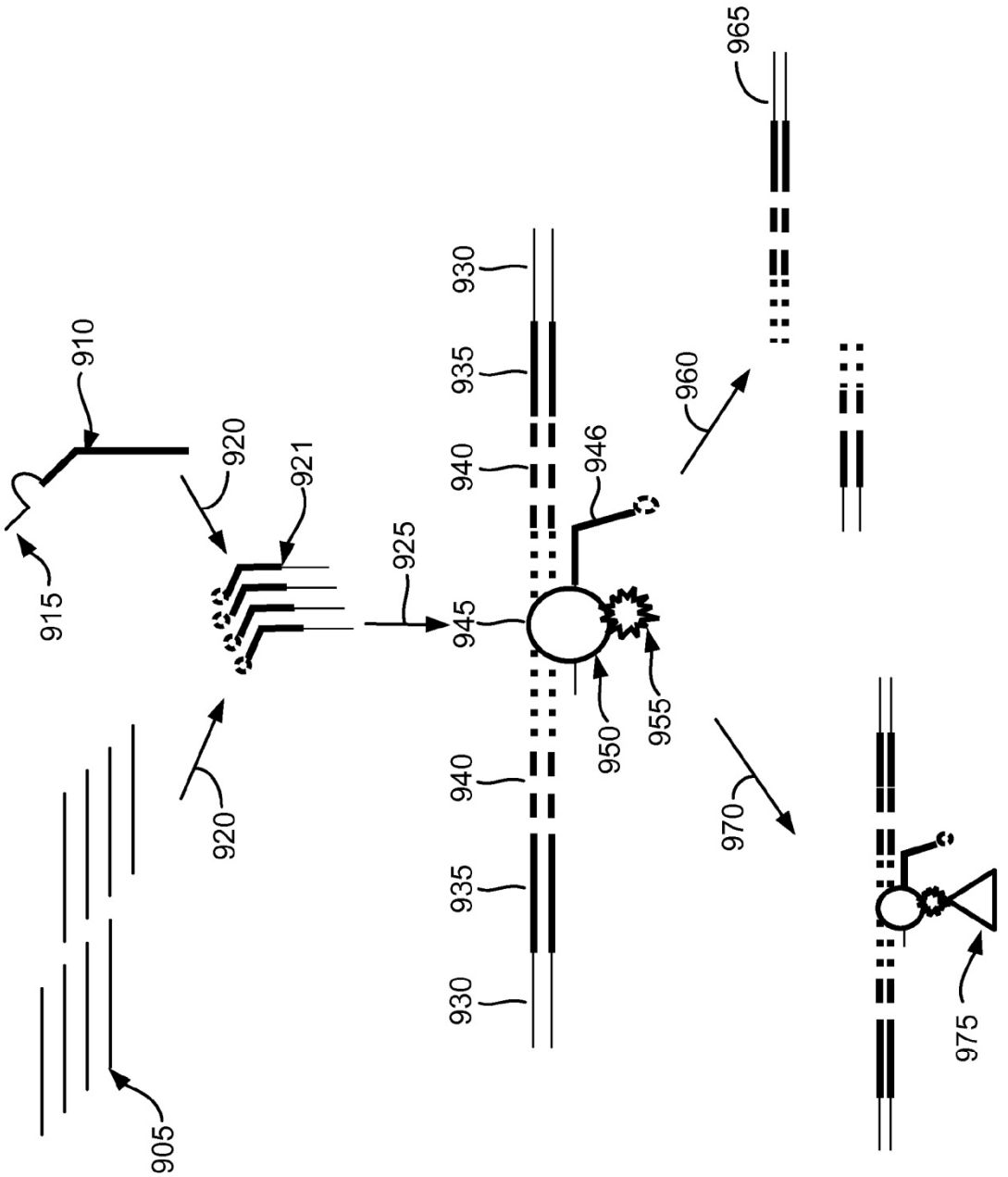

905
910
915
920
921
925
930
935
940
945
946
950
955
960
965
970
975

FIG. 9

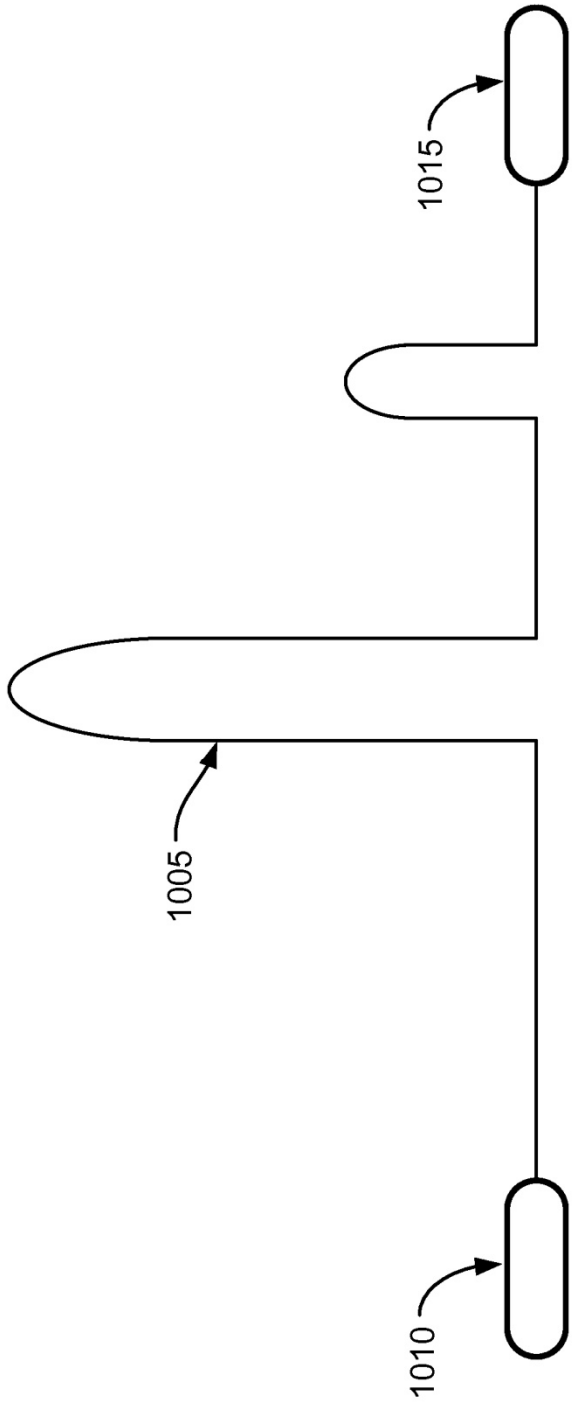


FIG. 10

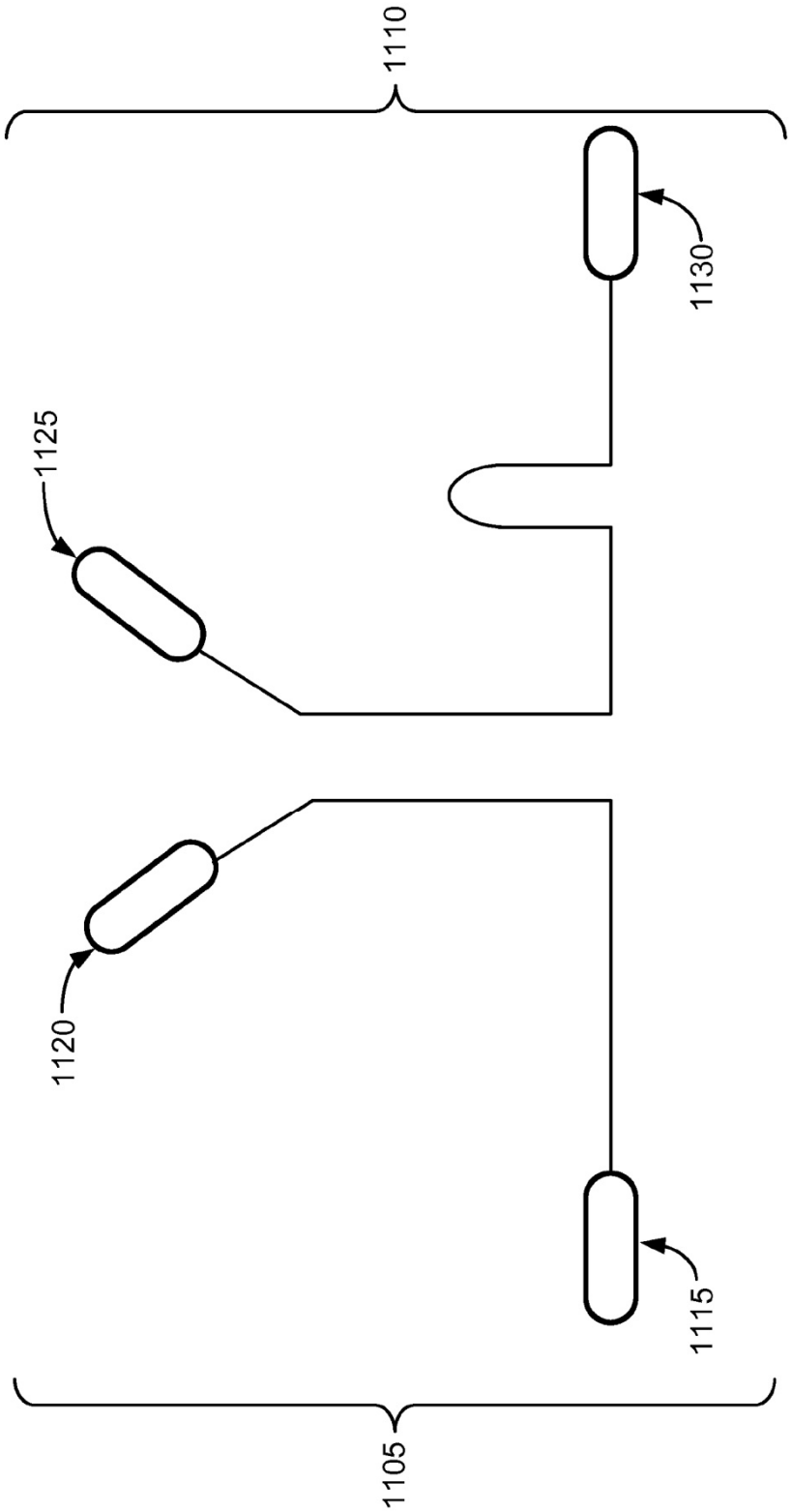


FIG. 11

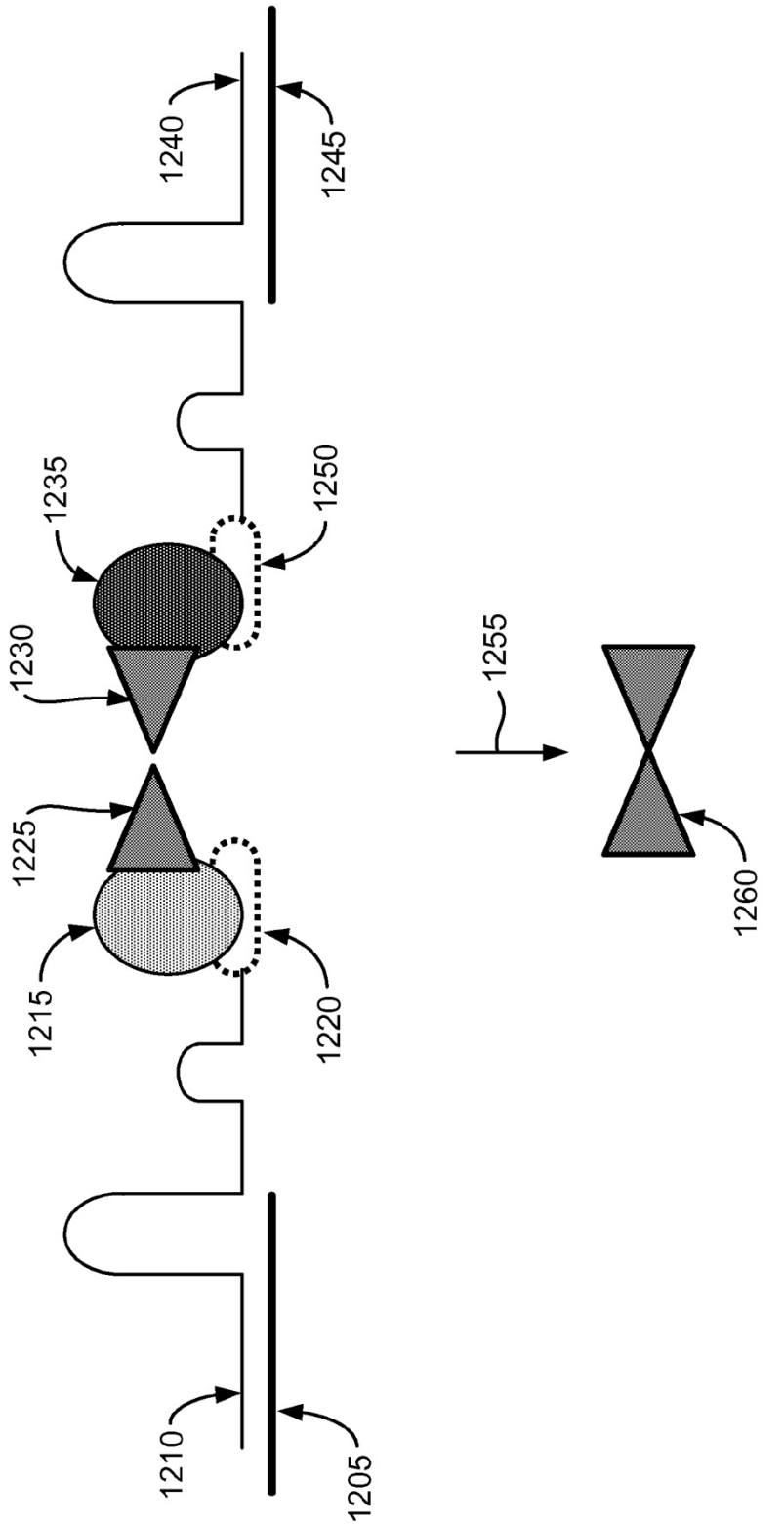
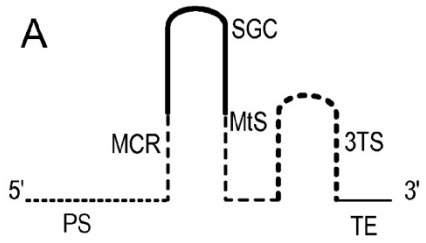
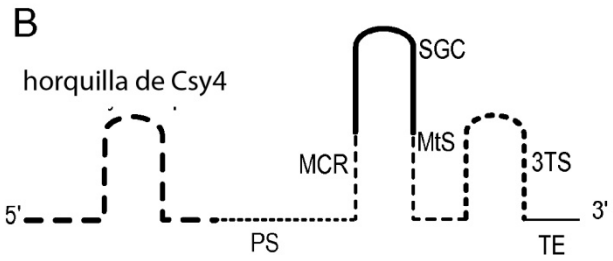


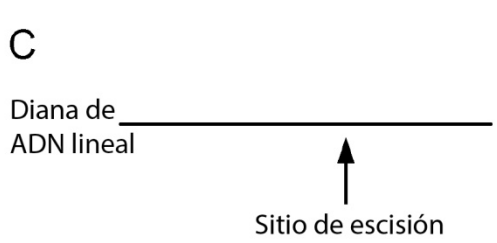
FIG. 12



ARNgs sin marcar



ARNgs marcado



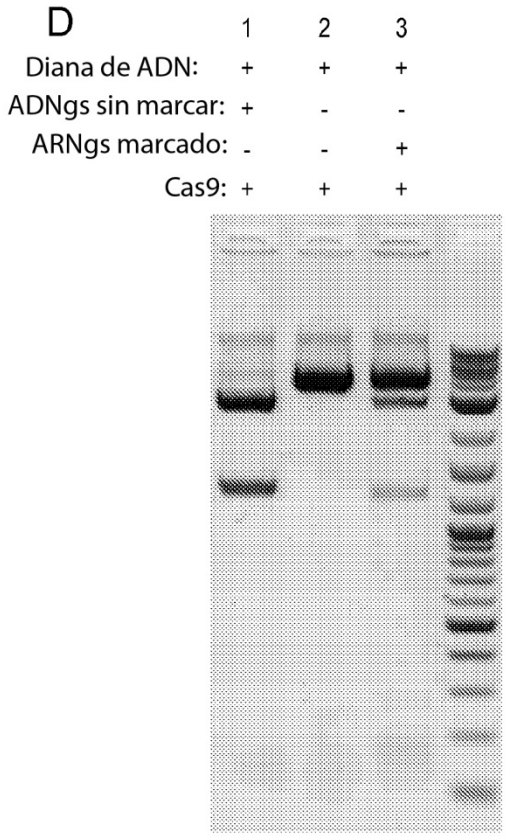


FIG. 13

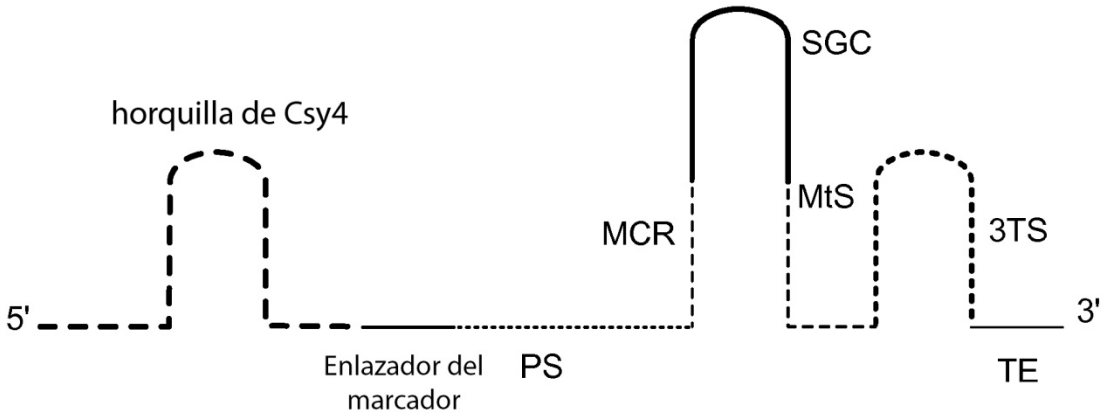

SGC
horquilla de Csy4
MtS
MCR
3TS
5'
3'
Enlazador del marcador
PS
TE

FIG. 14

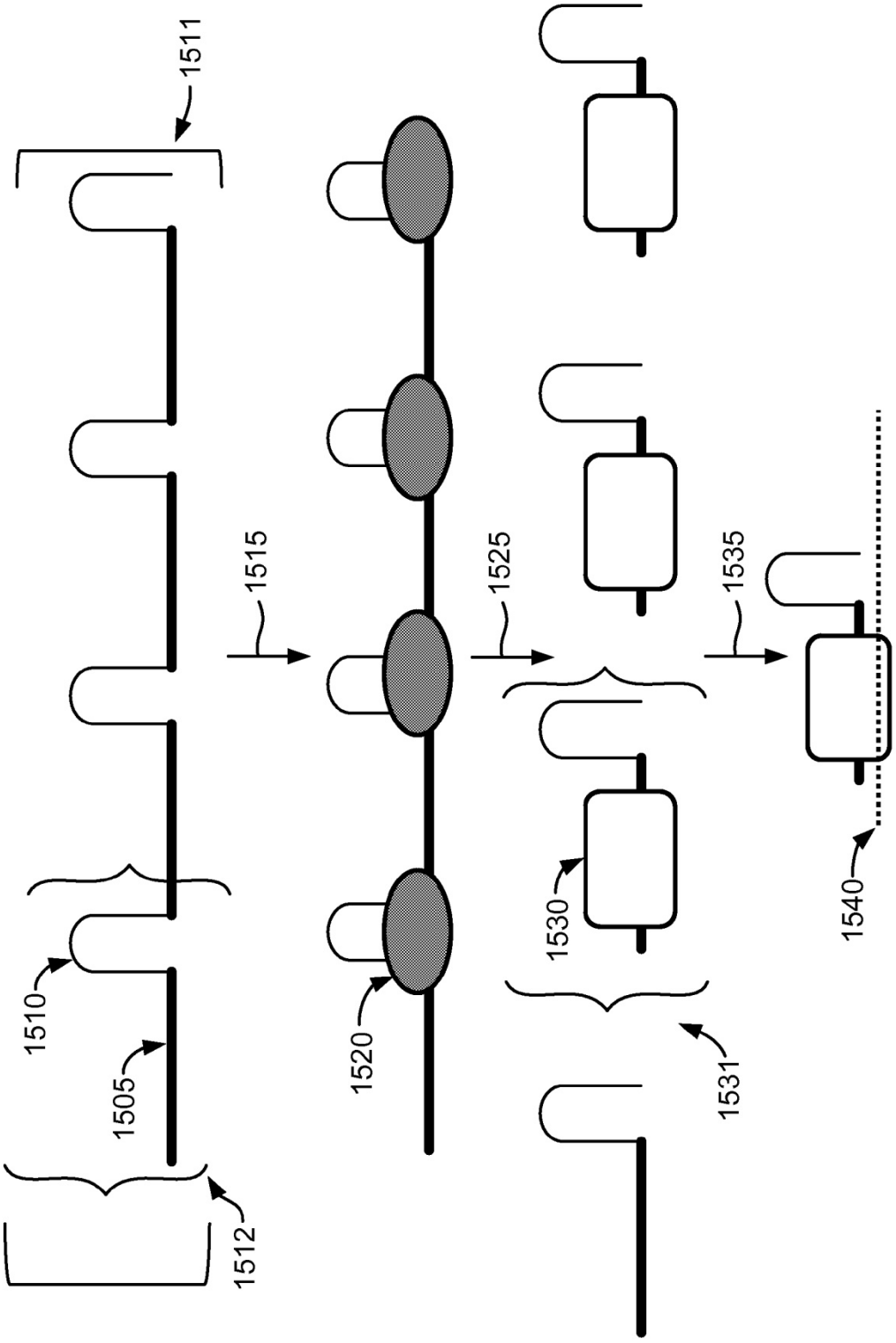


FIG. 15

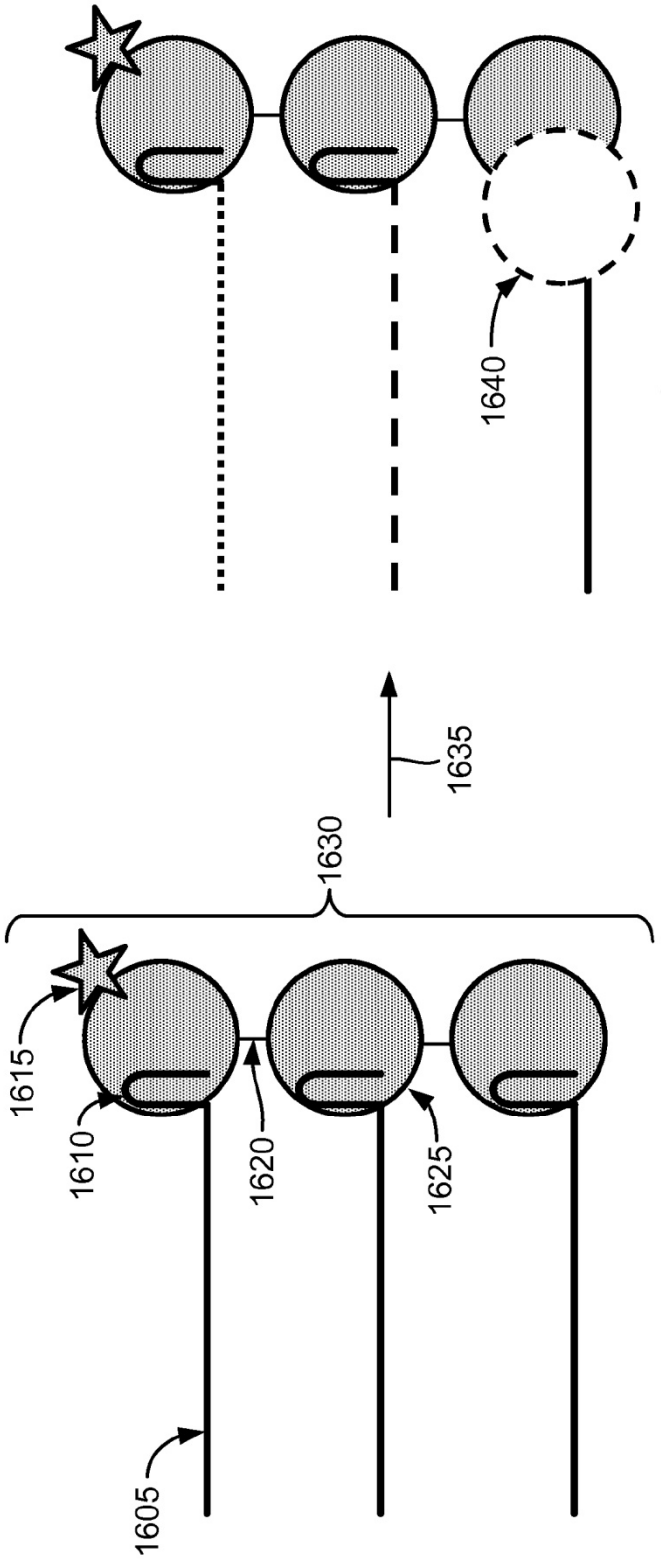


FIG. 16

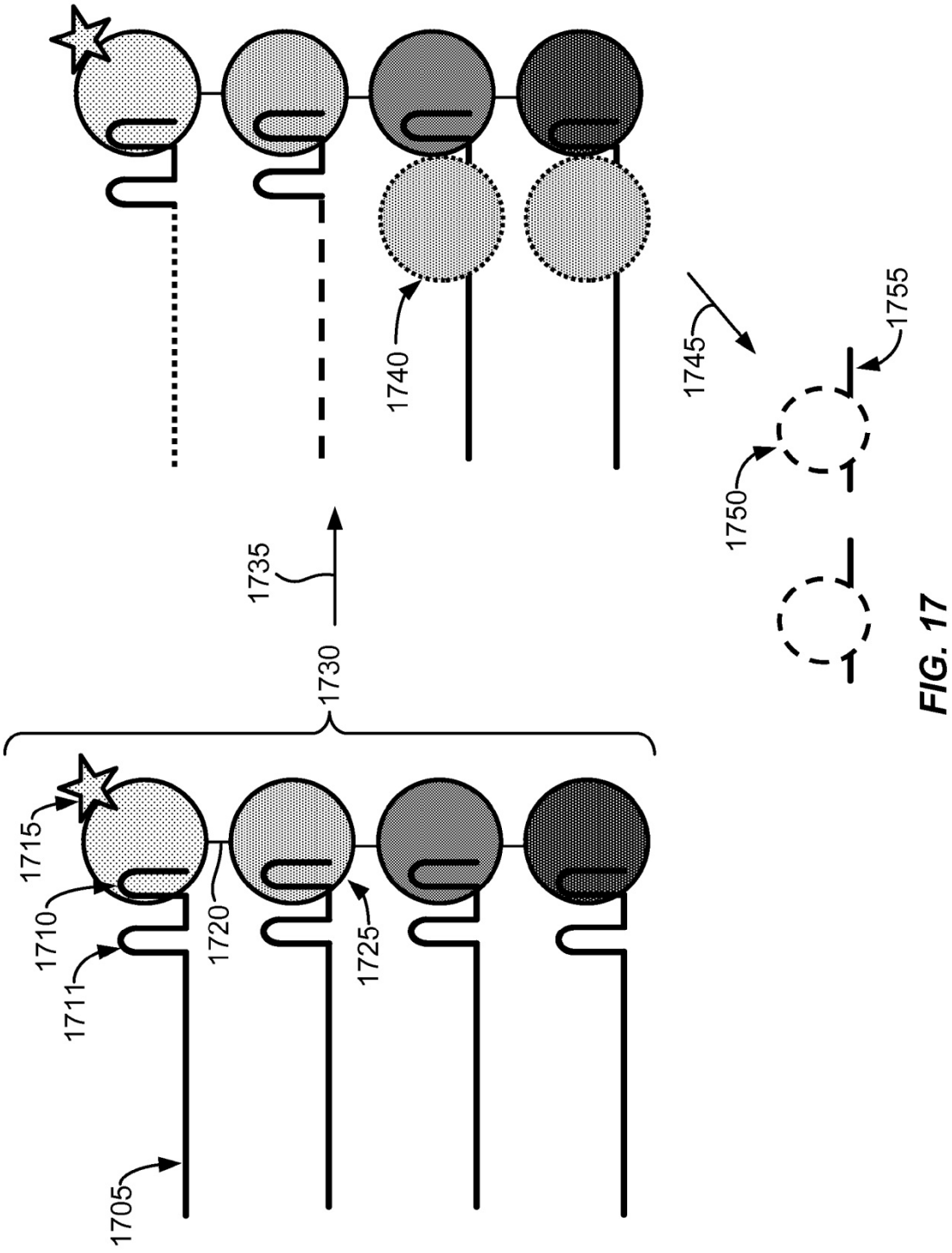


FIG. 17

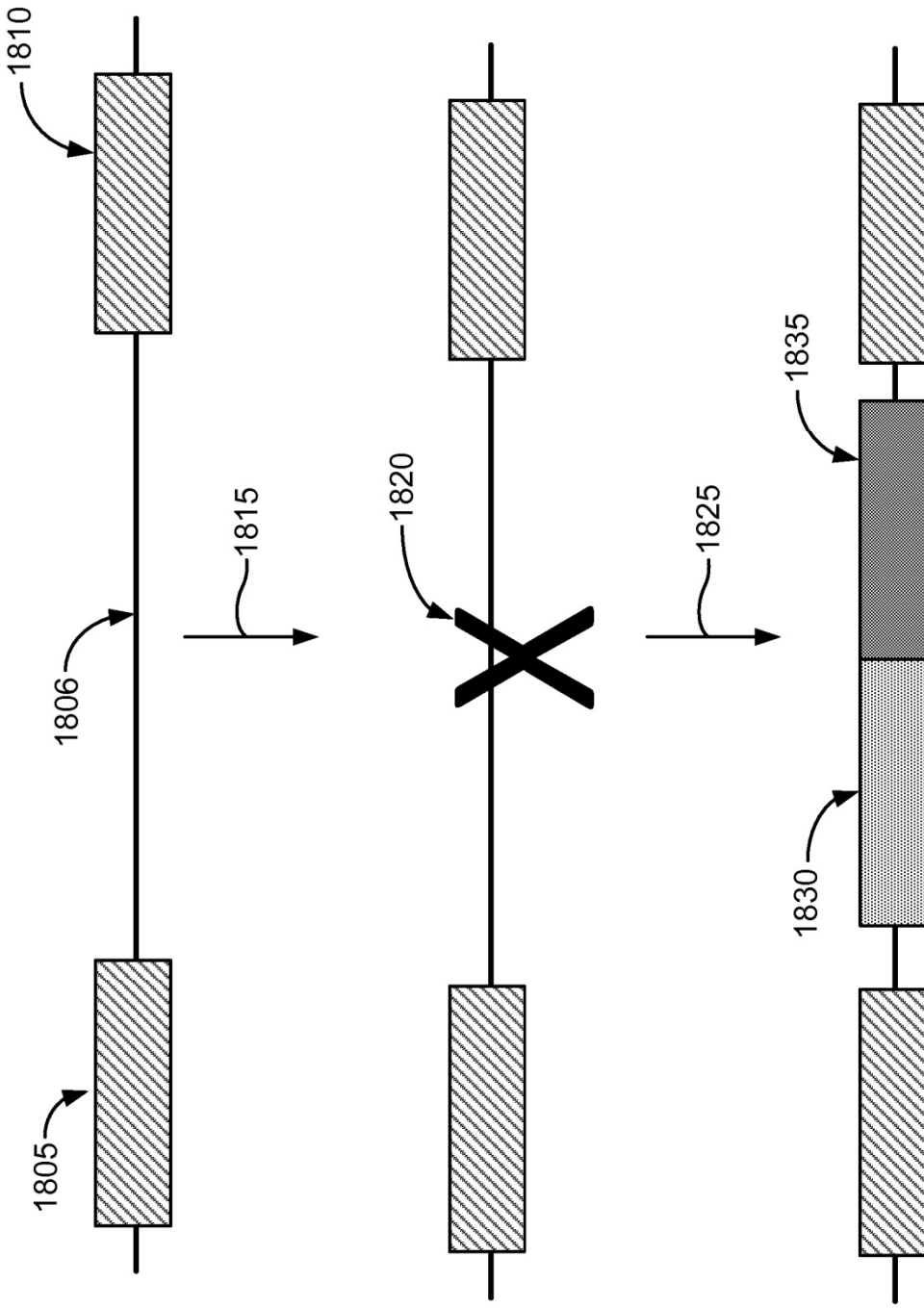


FIG. 18

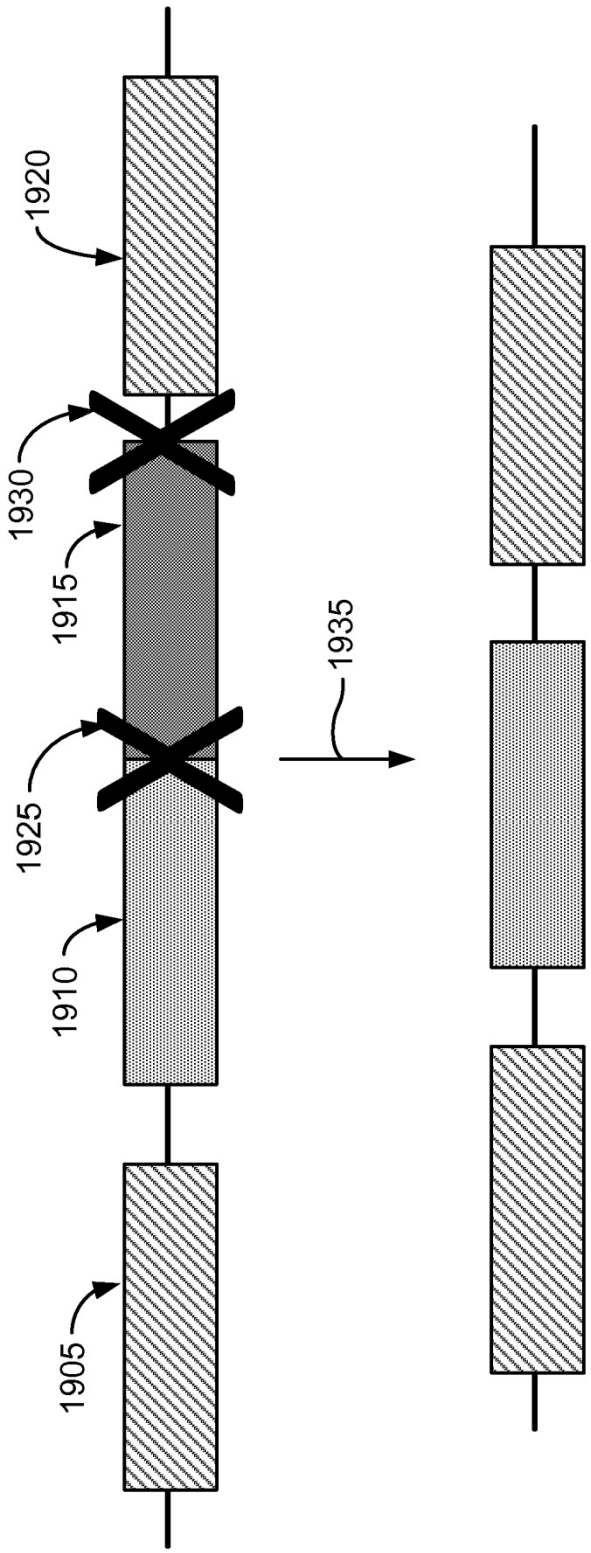


FIG. 19

5′ –GUUUUAG–AGCUAUGCUGUUUUGAAUGGUCCCAAAAC–3′ – Secuencia de repetición de ARNcr
3′ –UAAAAUUGAACGAUACGACAAAACUUACCAAGGUUGUU–5′ – Secuencia de ARNcrtra

FIG. 20

FIG. 21

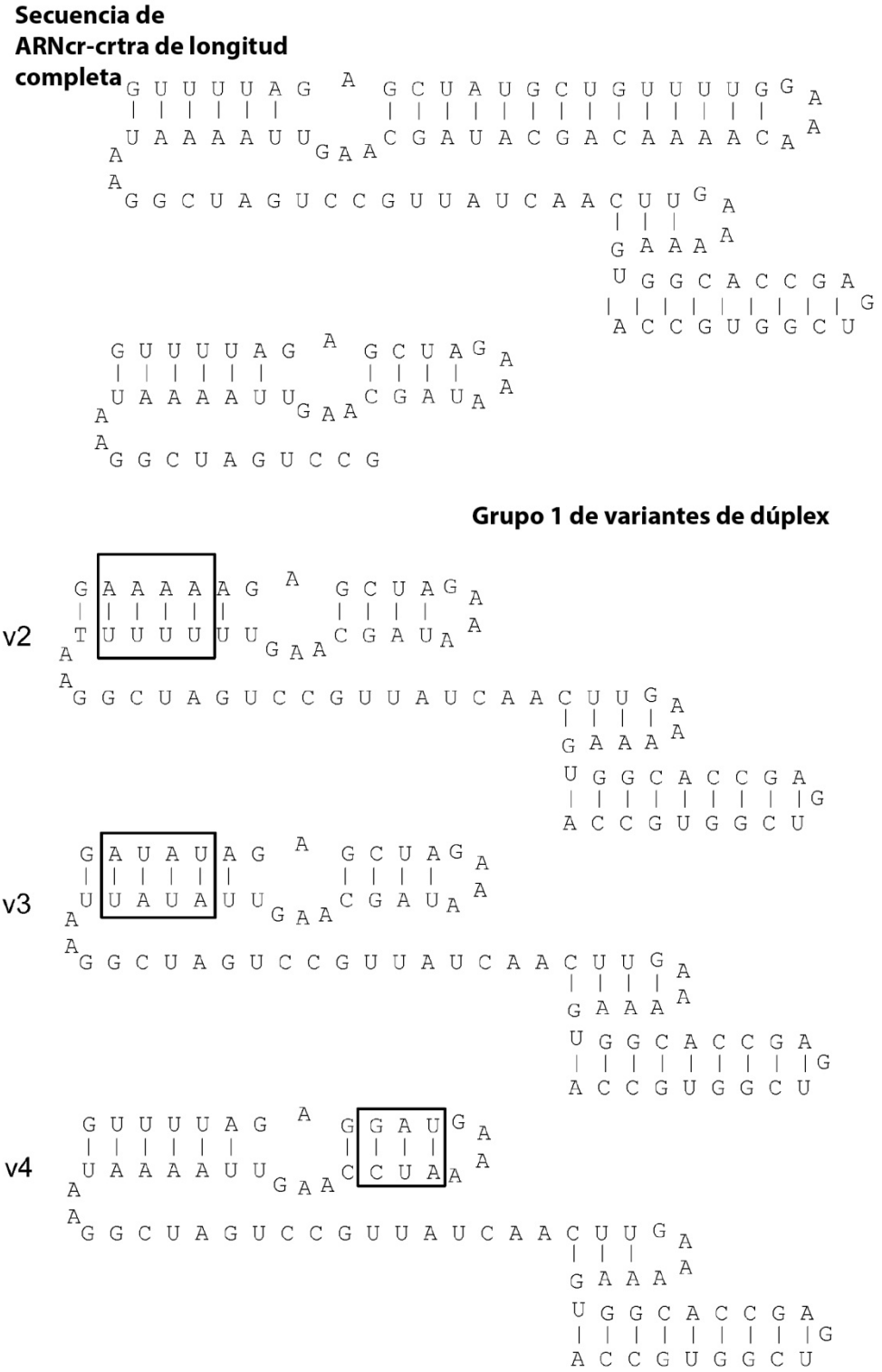

Secuencia de
ARNcr-crtra de longitud
completa
Grupo 1 de variantes de dúplex
v2
v3
v4

FIG. 22A

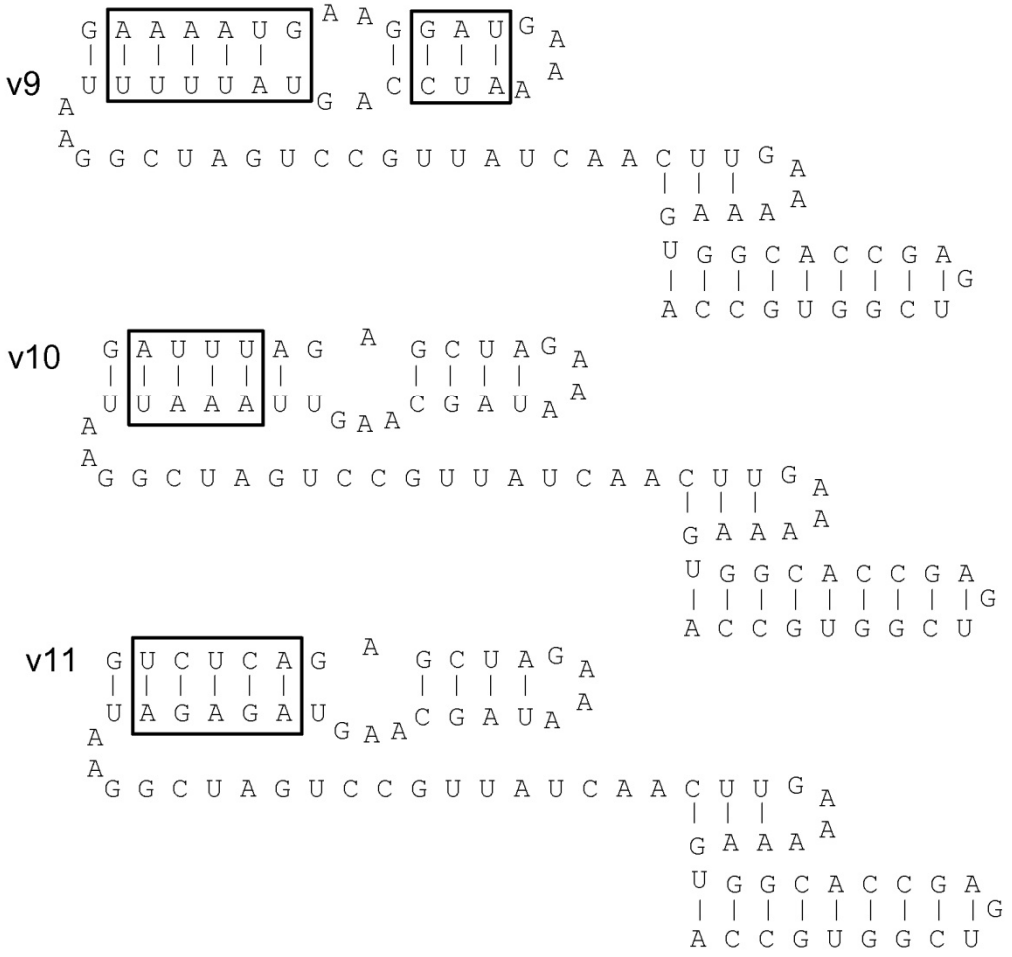


FIG. 22A
(Continuación)

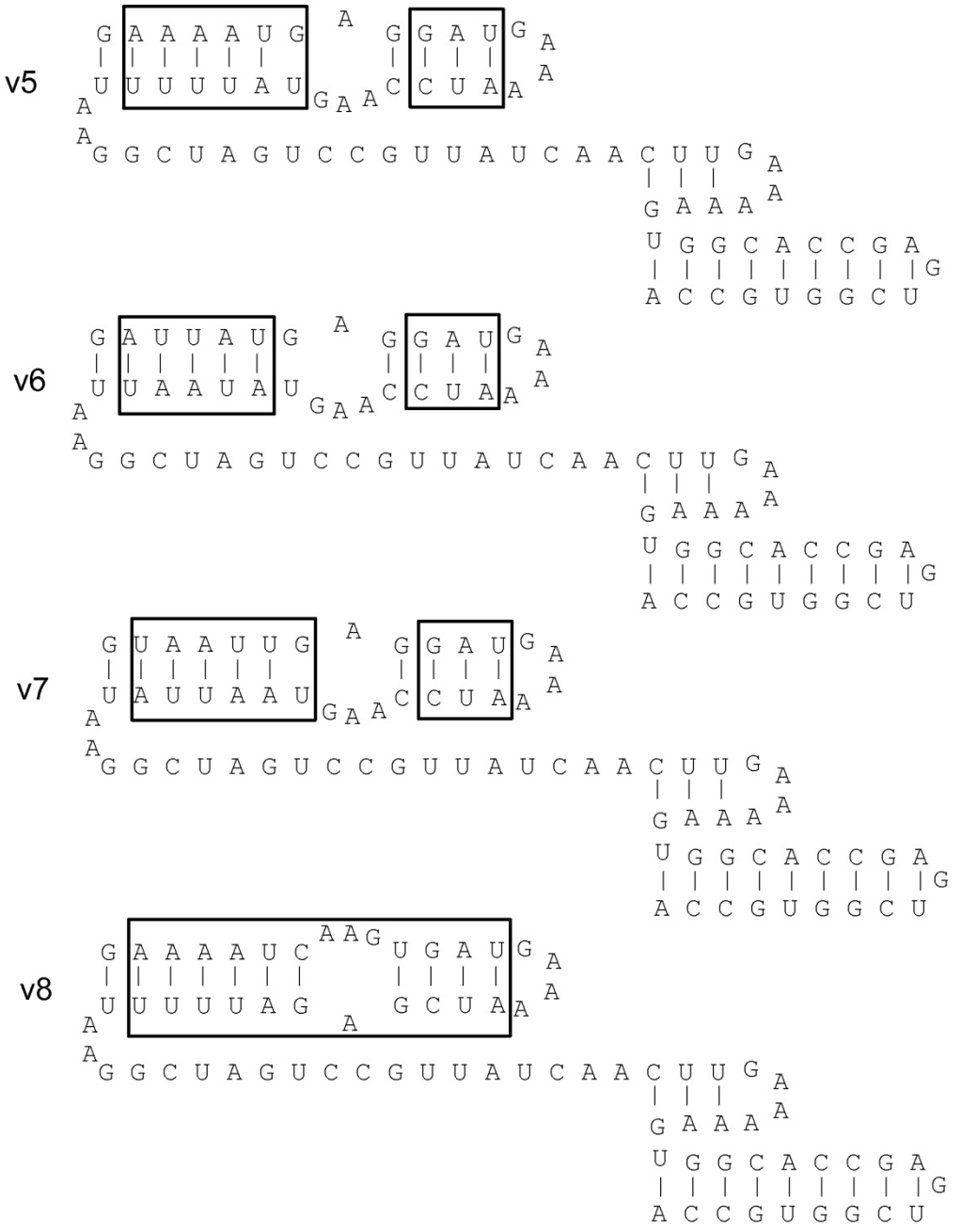


FIG. 22B

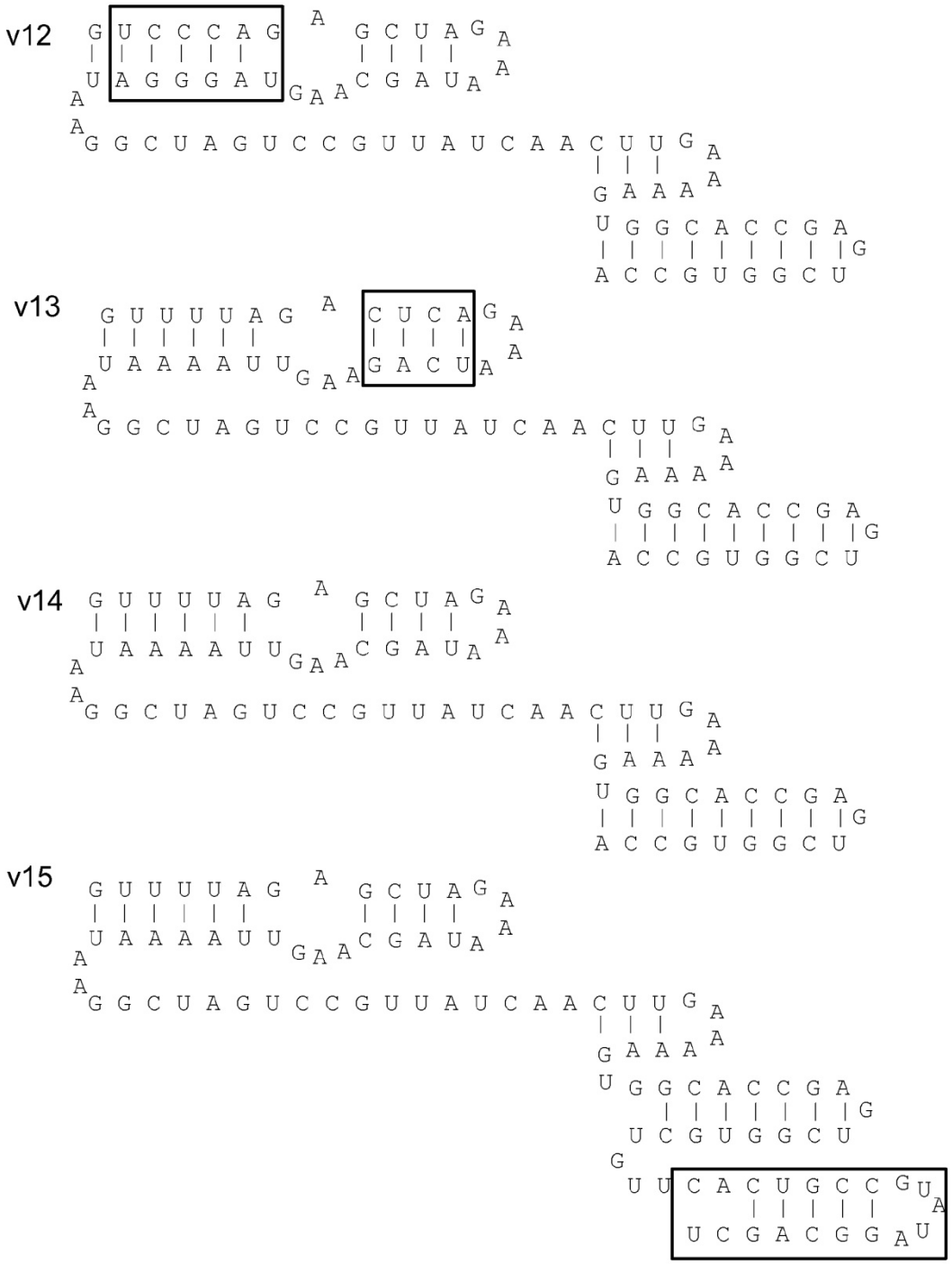


FIG. 22B
(Continuación)

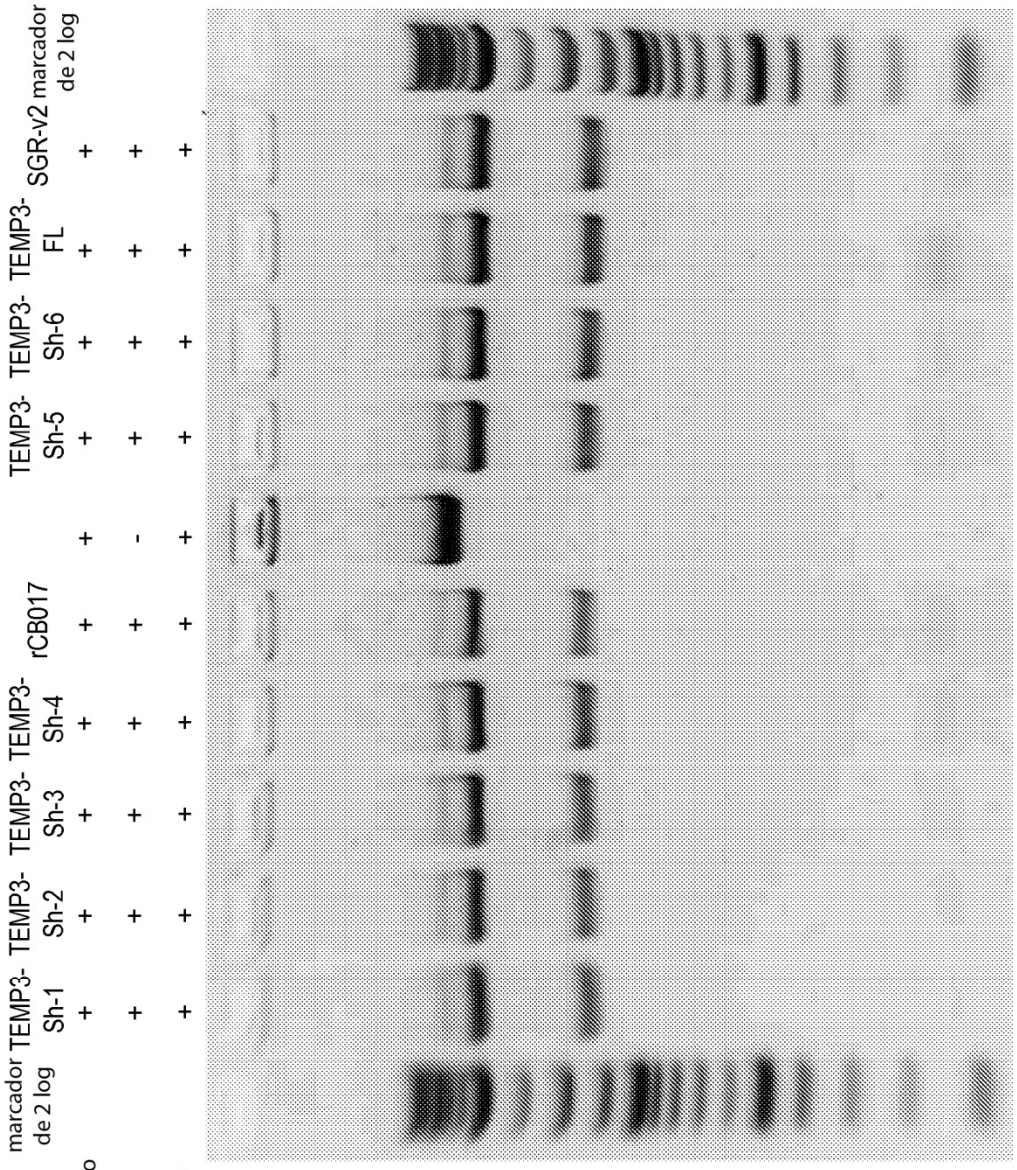

1
marcador de 2 log
2
TEMP3-Sh-1
3
TEMP3-Sh-2
4
TEMP3-Sh-3
5
TEMP3-Sh-4
6
rCB017
7
8
TEMP3-Sh-5
9
TEMP3-Sh-6
10
TEMP3-FL
11
SGR-v2
12
marcador de 2 log
plásmido
ARNgs
Cas9

FIG. 23A

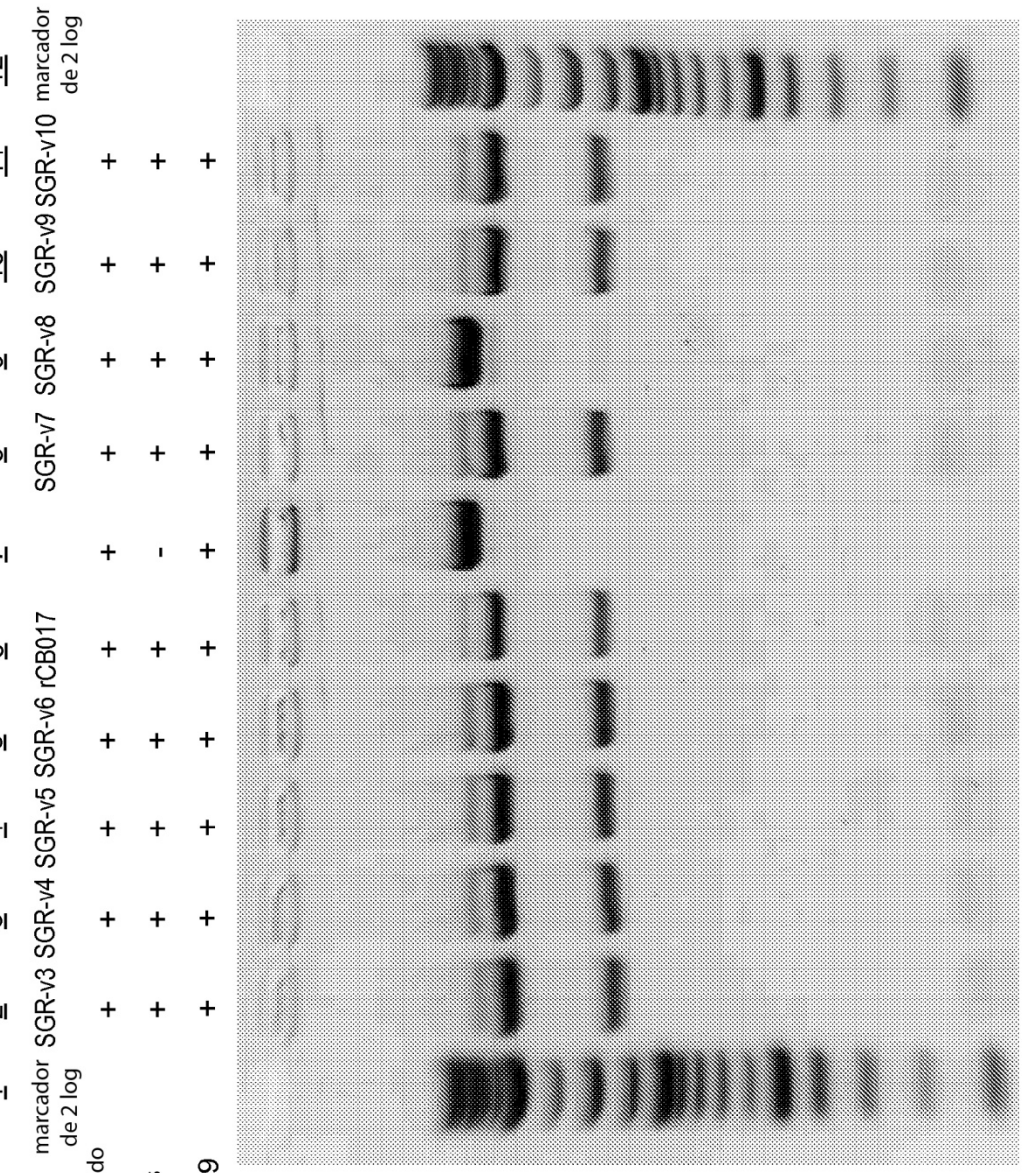

1 2 3 4 5 6 7 8 9 10 11 12
marcador de 2 log
SGR-v3
SGR-v4
SGR-v5
SGR-v6
rCB017
SGR-v7
SGR-v8
SGR-v9
SGR-v10
marcador de 2 log
plásmido
ARNgs
Cas9

FIG. 23B

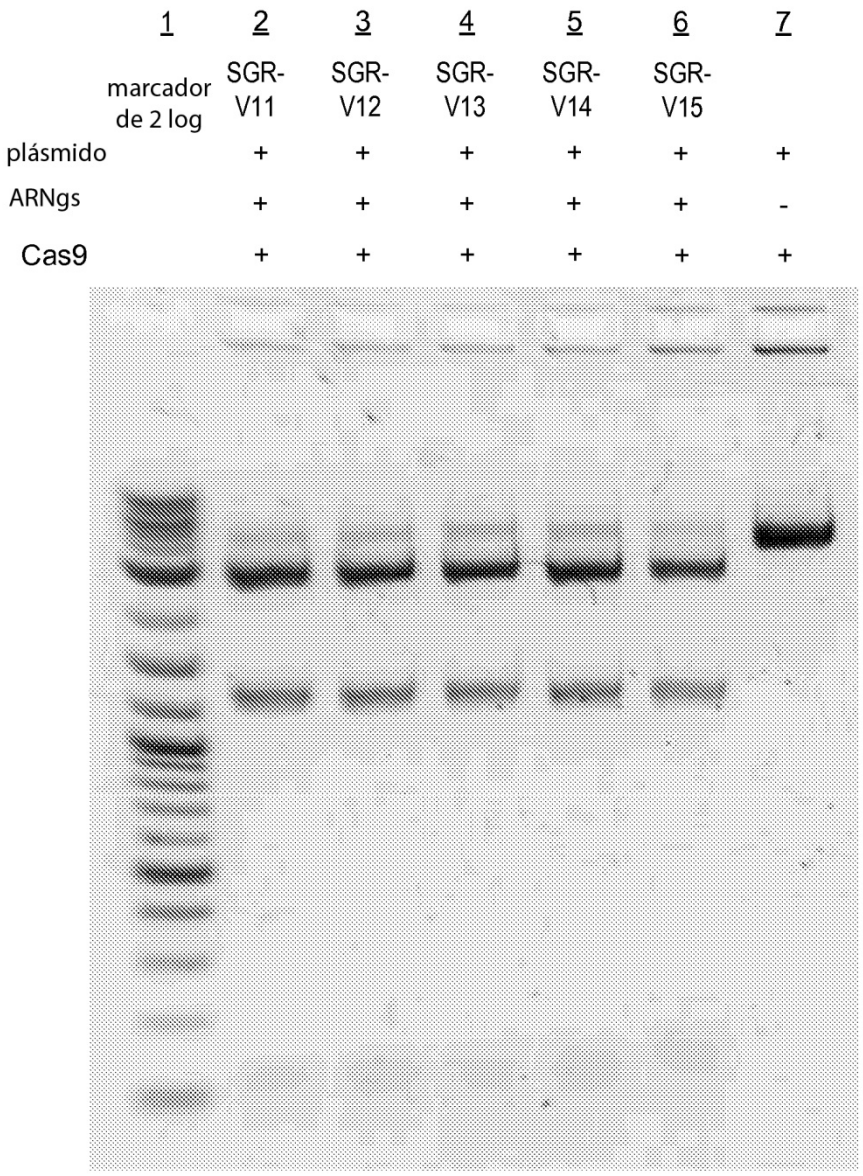

1
2
3
4
5
6
7
marcador de 2 log
SGR-V11
SGR-V12
SGR-V13
SGR-V14
SGR-V15
plásmido + + + + + +
ARNgs + + + + + -
Cas9 + + + + + +

FIG. 23C

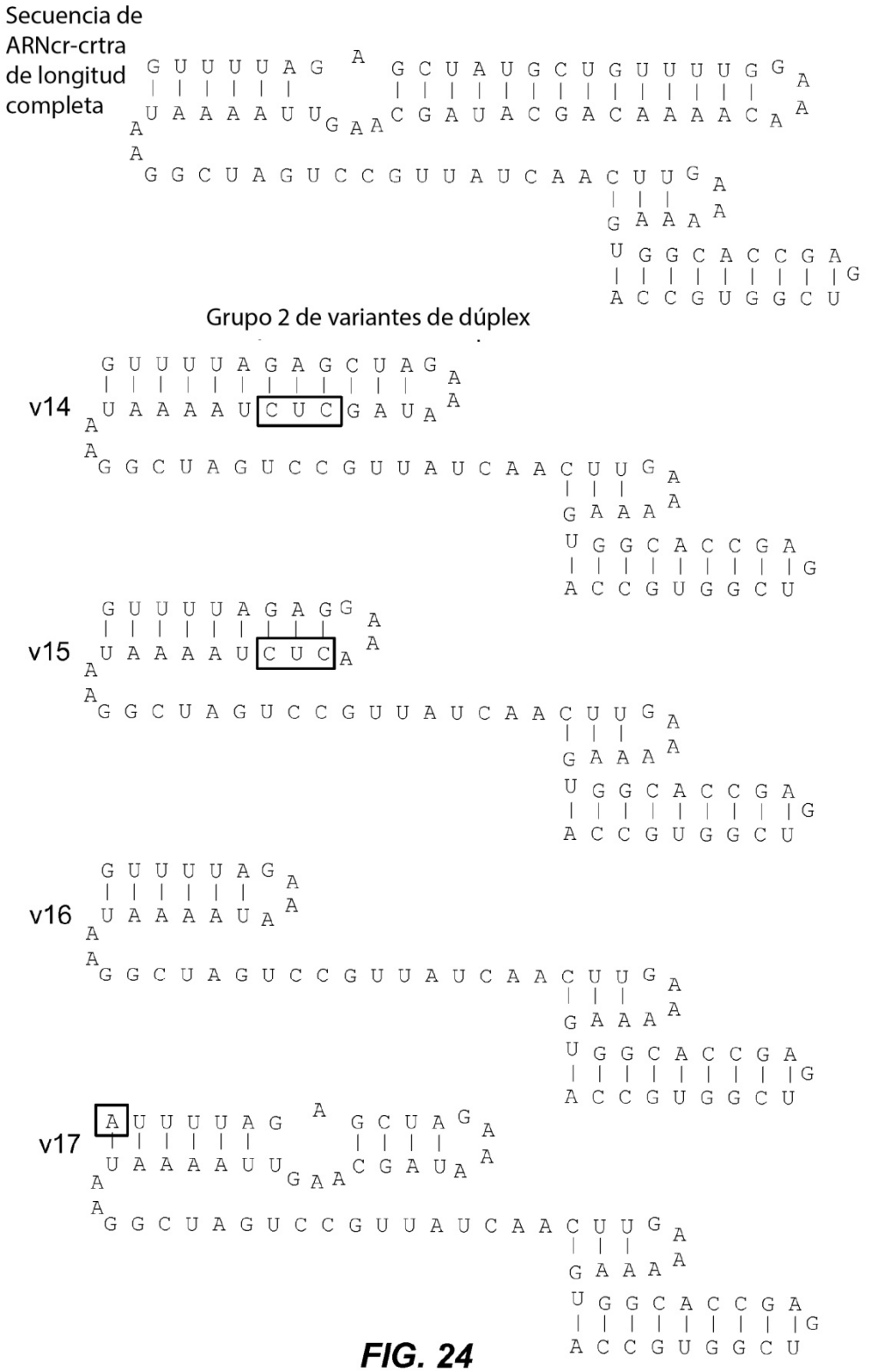

Secuencia de ARNcr-crtra de longitud completa
Grupo 2 de variantes de dúplex
v14
v15
v16
v17

FIG. 24

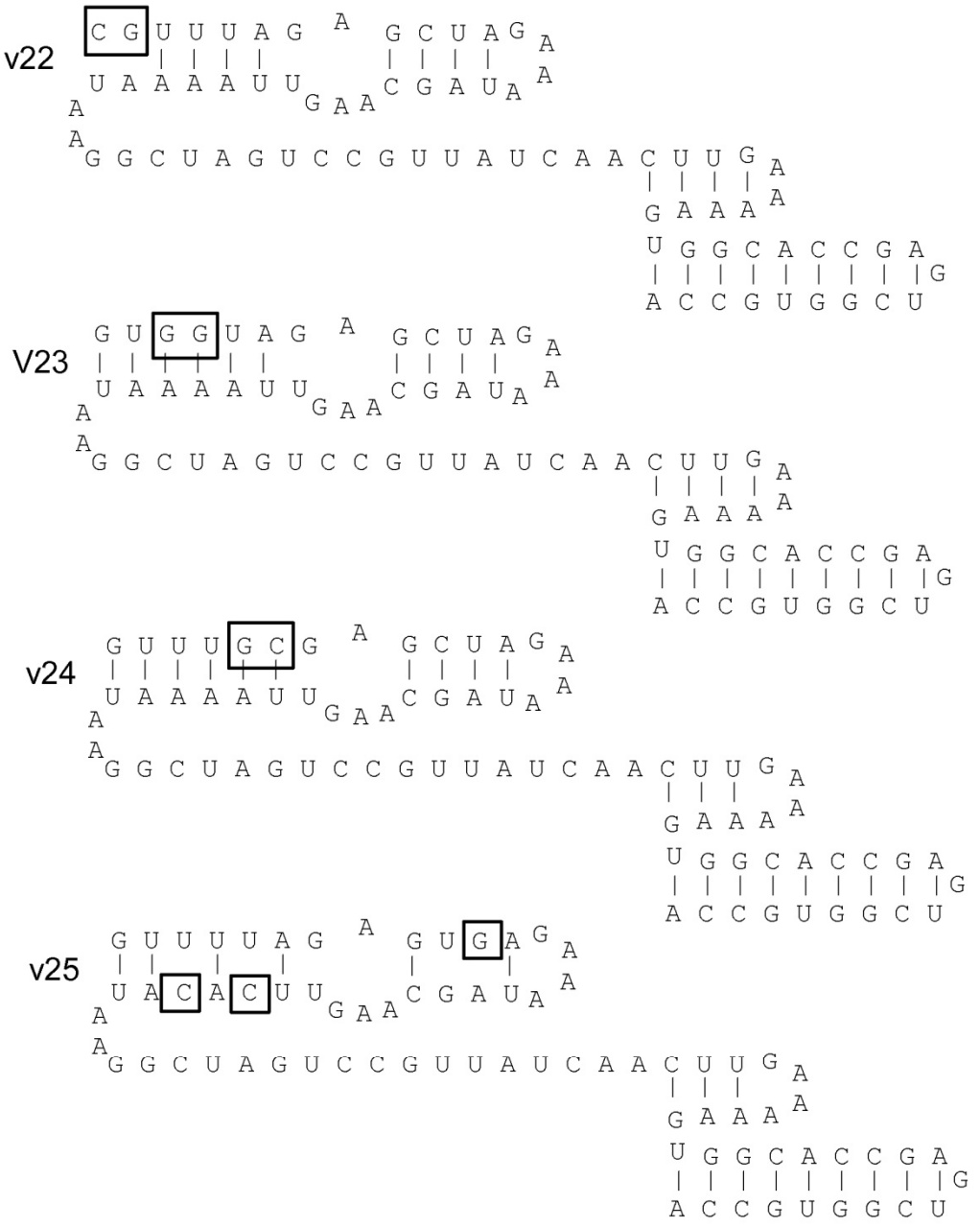


FIG. 24
(Continuación)

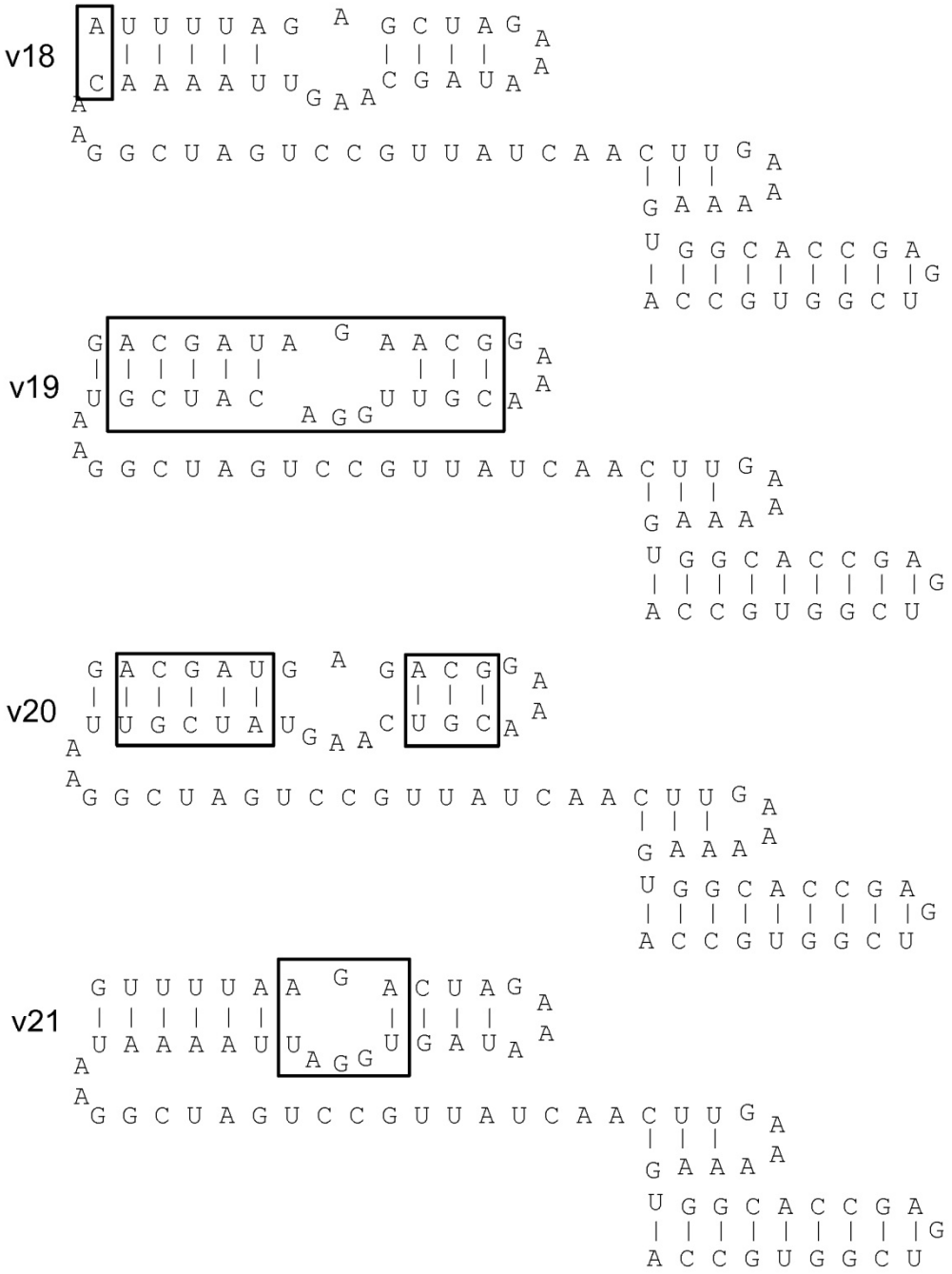


FIG. 25

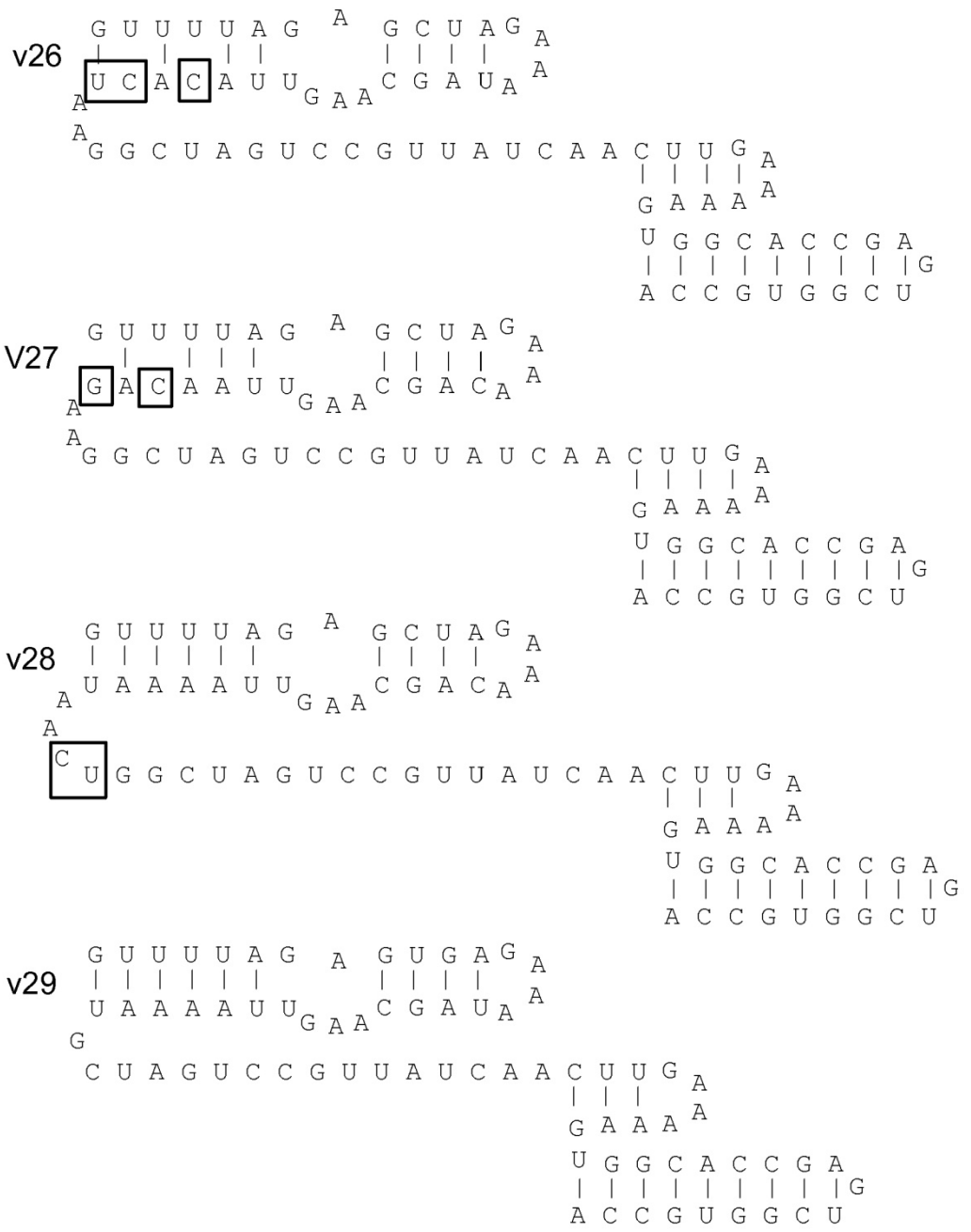


FIG. 25B
(Continuación)

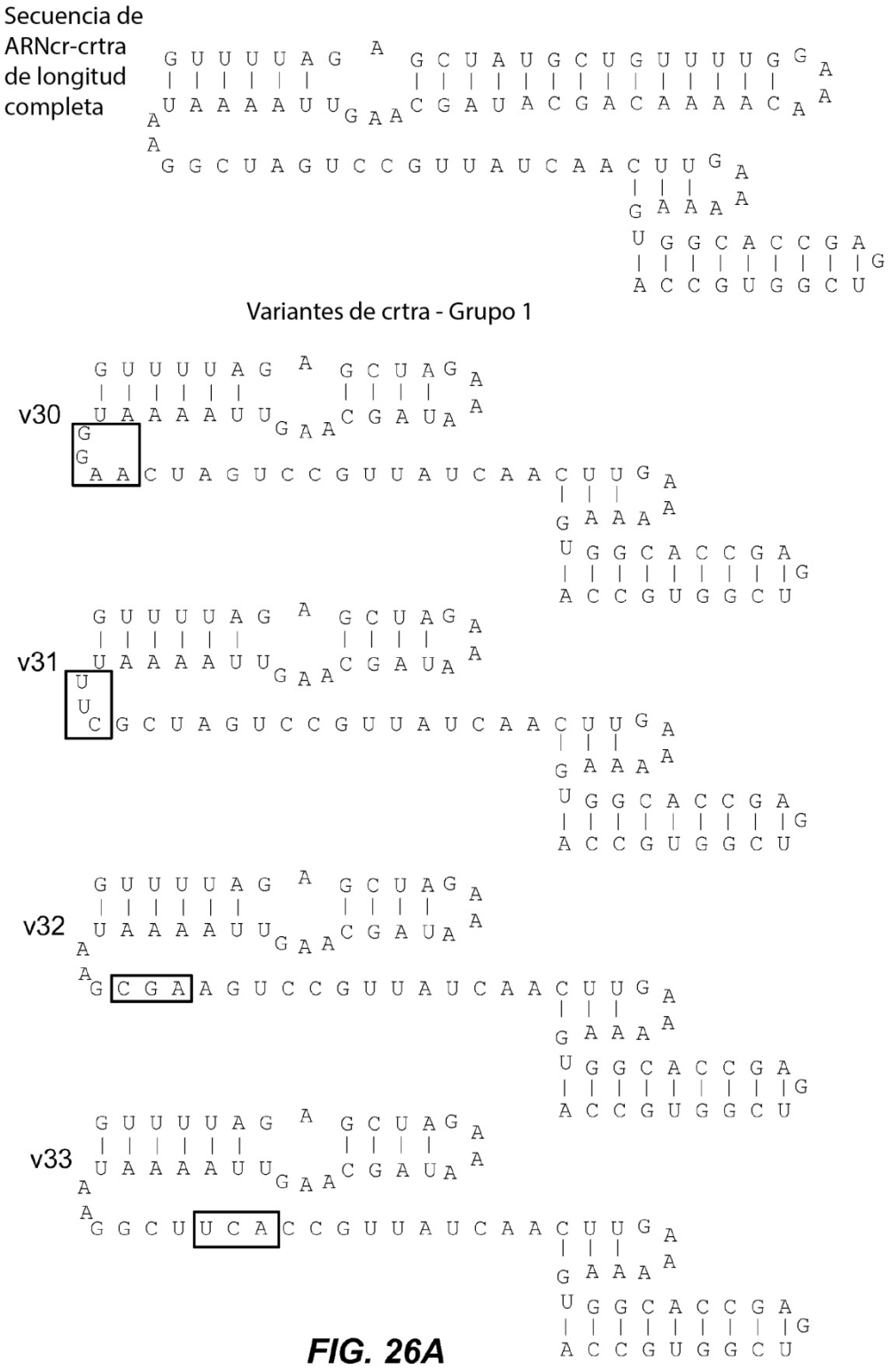

Secuencia de ARNcr-crtra de longitud completa
Variantes de crtra - Grupo 1
v30
v31
v32
v33

FIG. 26A

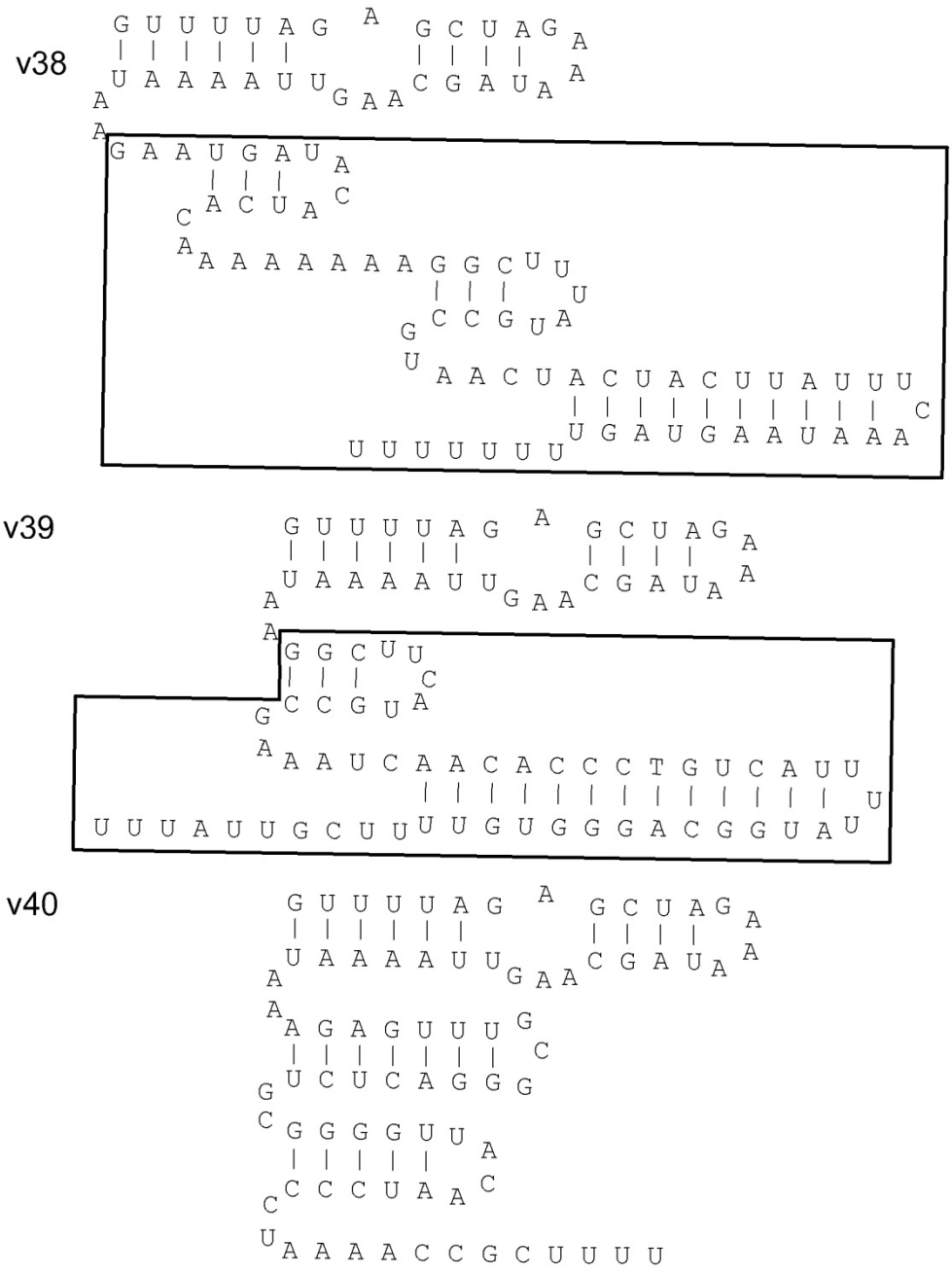


FIG. 26A
(Continuación)

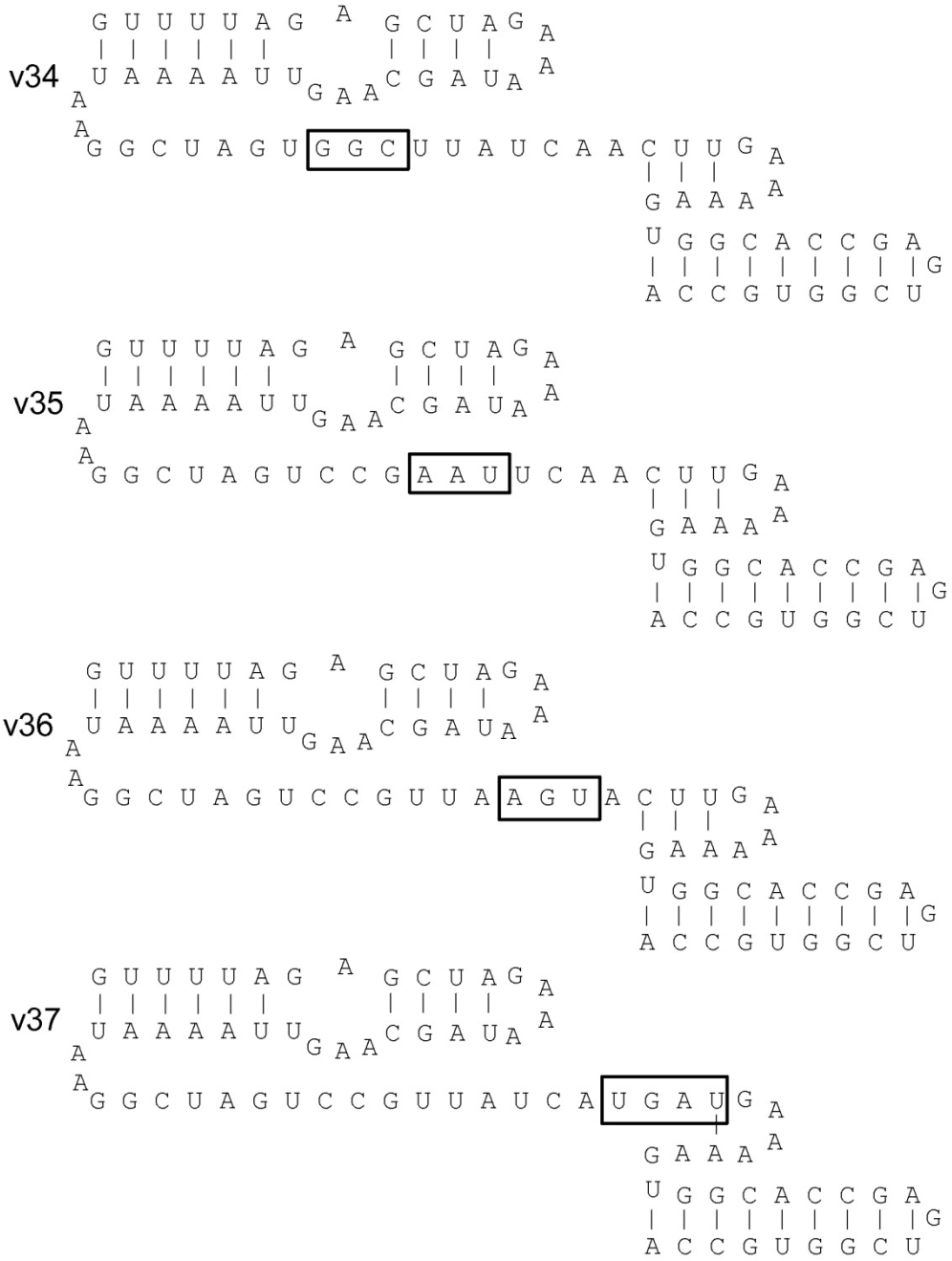


FIG. 26B

v41

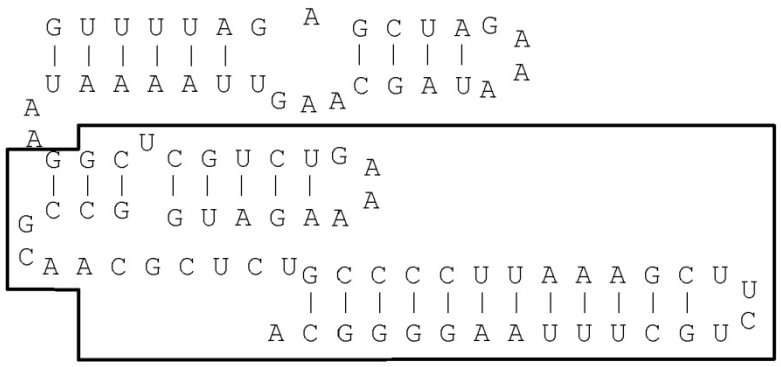


FIG. 26B
(Continuación)

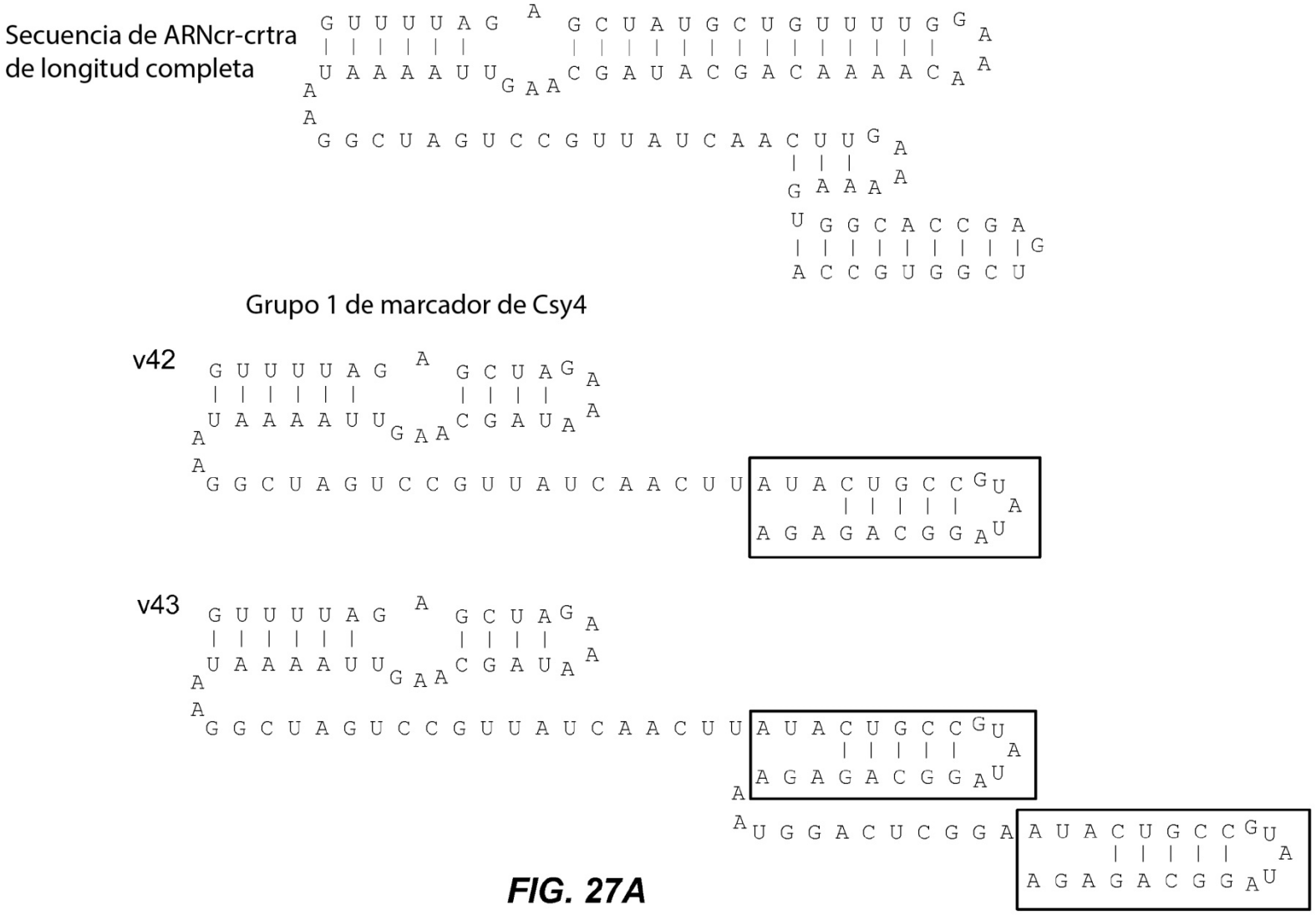

Secuencia de ARNcr-crtra de longitud completa
Grupo 1 de marcador de Csy4
v42
v43

FIG. 27A

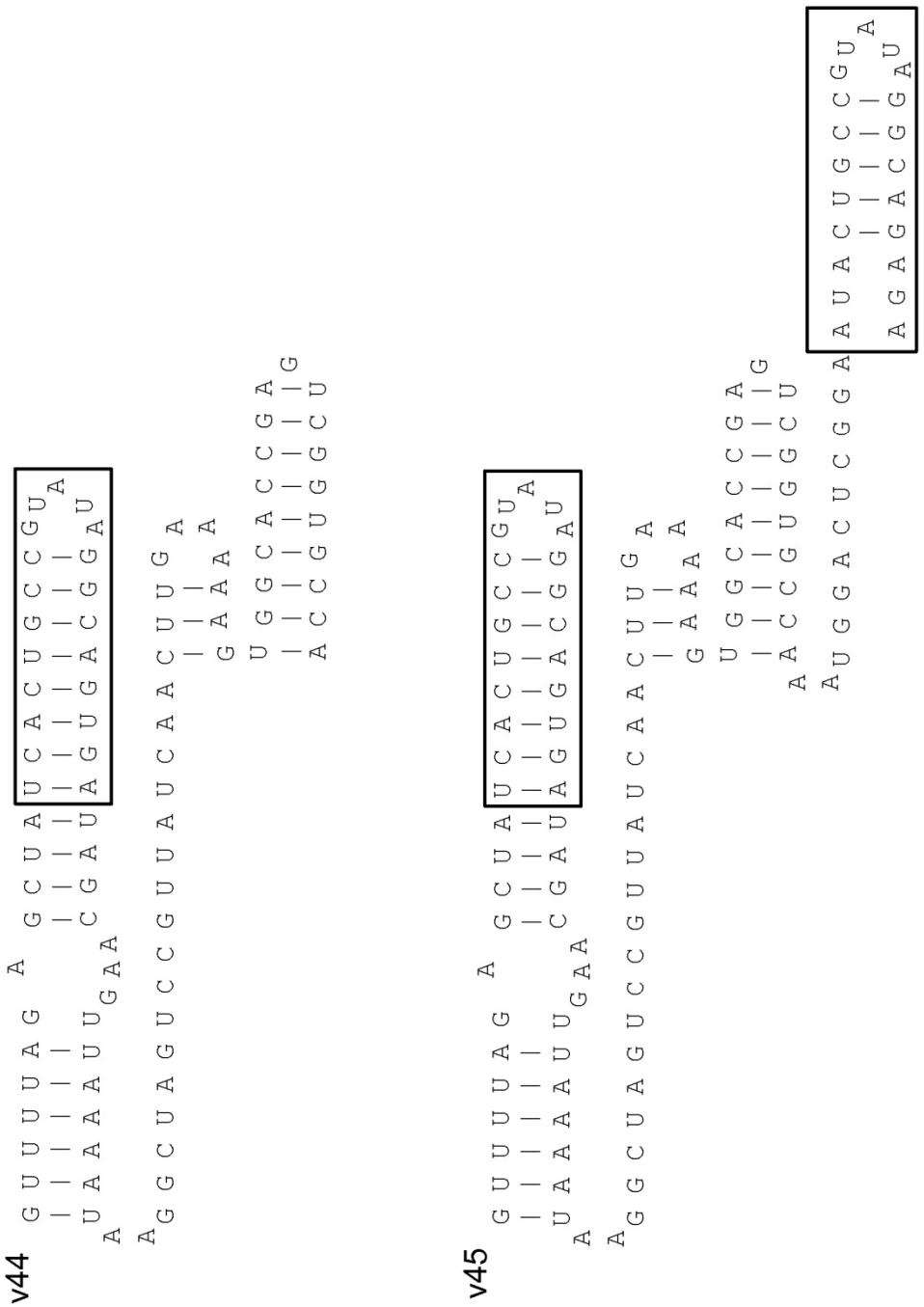


FIG. 27B

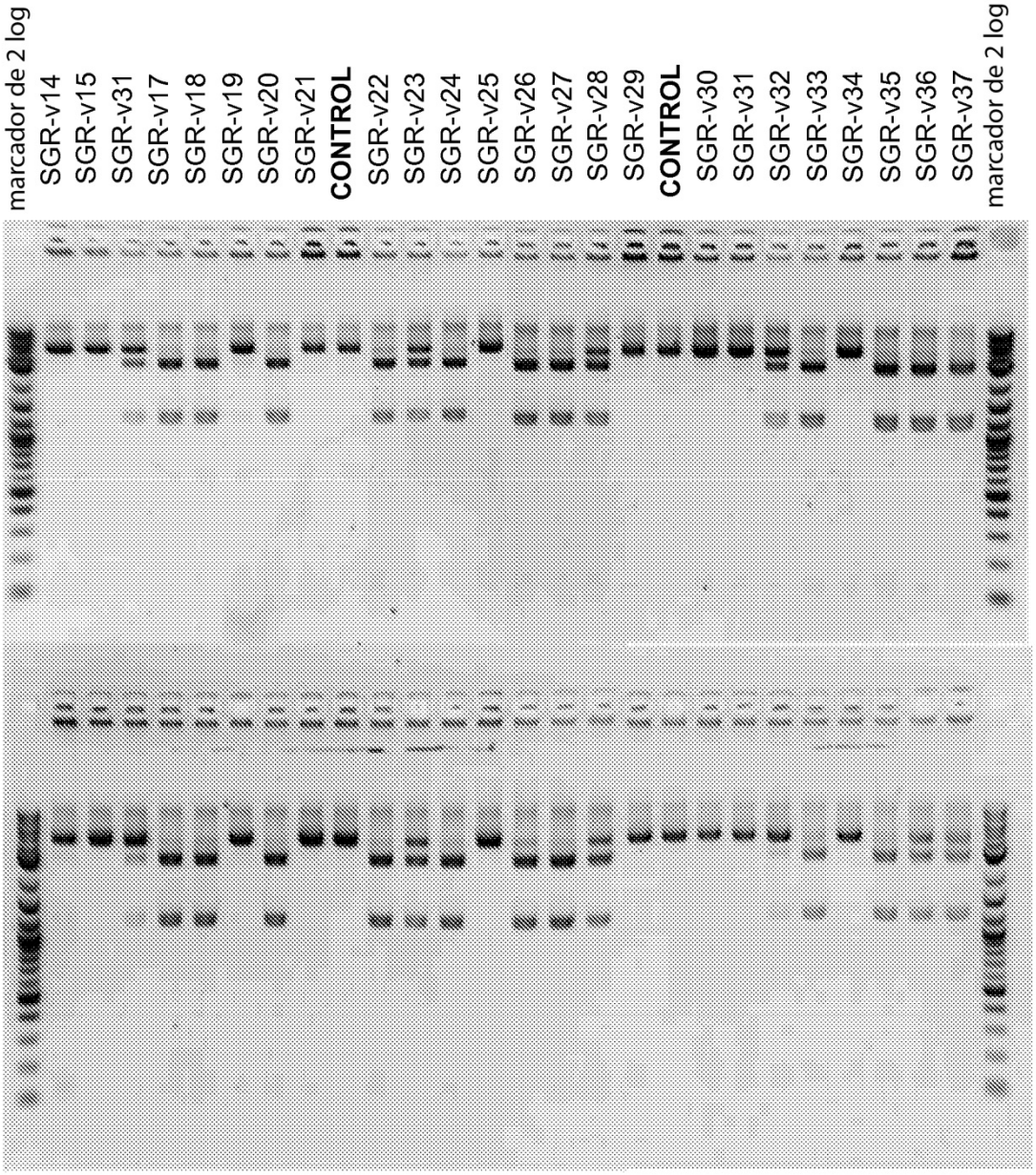

marcador de 2 log
SGR-v14
SGR-v15
SGR-v31
SGR-v17
SGR-v18
SGR-v19
SGR-v20
SGR-v21
CONTROL
SGR-v22
SGR-v23
SGR-v24
SGR-v25
SGR-v26
SGR-v27
SGR-v28
SGR-v29
CONTROL
SGR-v30
SGR-v31
SGR-v32
SGR-v33
SGR-v34
SGR-v35
SGR-v36
SGR-v37
marcador de 2 log

FIG. 28

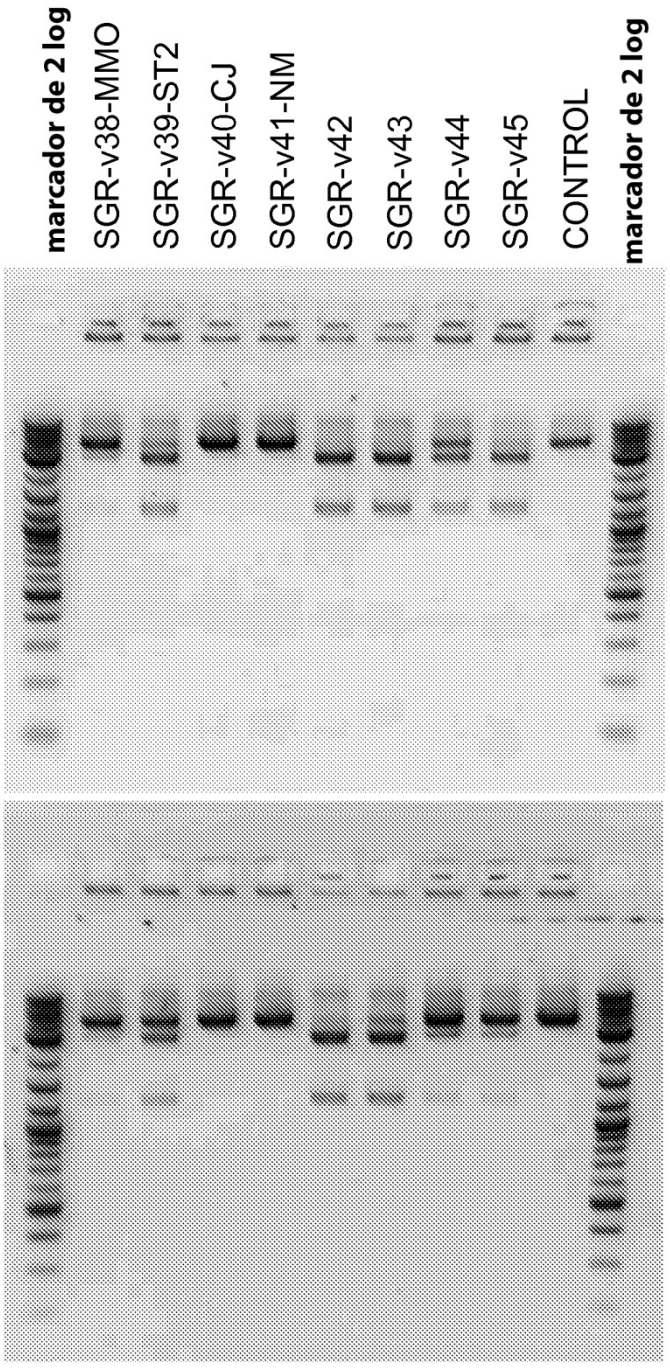

marcador de 2 log
SGR-v38-MMO
SGR-v39-ST2
SGR-v40-CJ
SGR-v41-NM
SGR-v42
SGR-v43
SGR-v44
SGR-v45
CONTROL
marcador de 2 log

FIG. 29

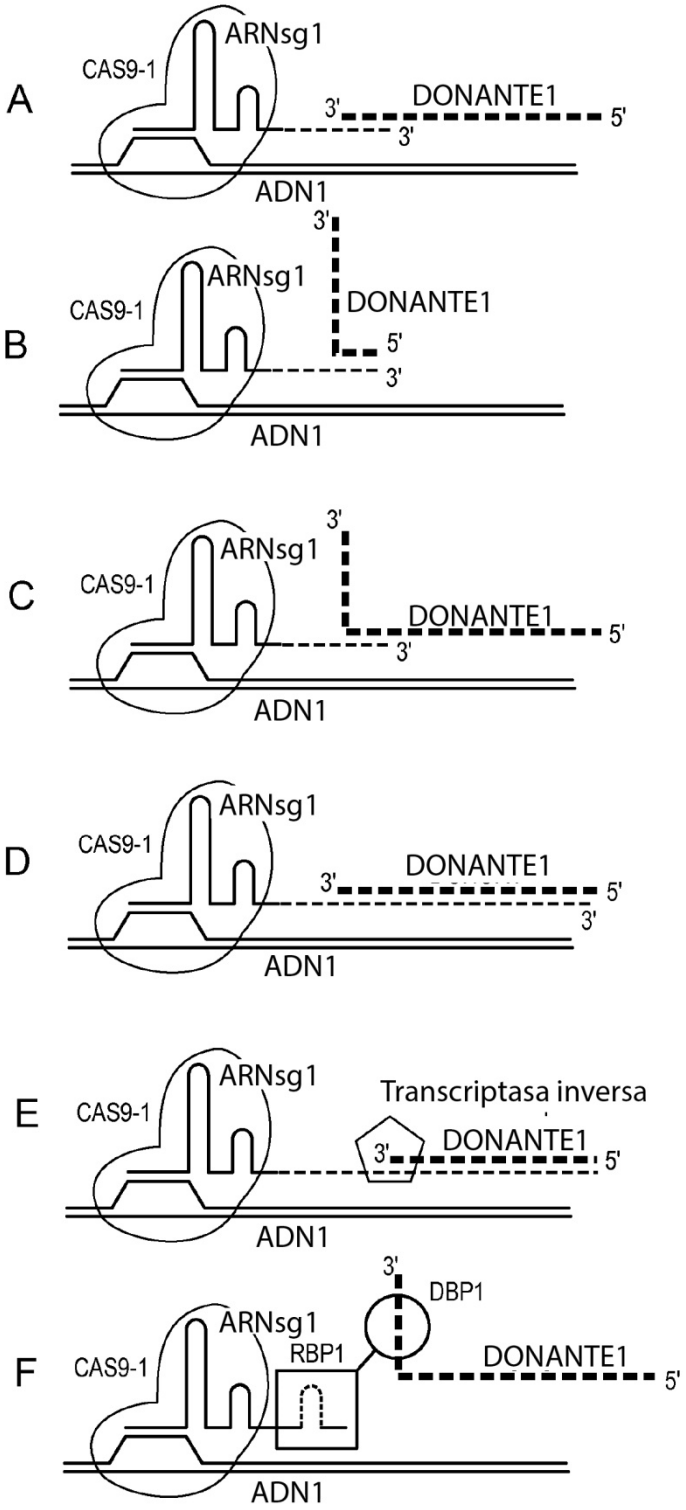

A
CAS9-1
ARNsg1
3'
DONANTE1
5'
3'
ADN1
B
3'
CAS9-1
ARNsg1
DONANTE1
5'
3'
ADN1
C
3'
CAS9-1
ARNsg1
DONANTE1
5'
3'
ADN1
D
CAS9-1
ARNsg1
3'
DONANTE1
5'
3'
ADN1
E
CAS9-1
ARNsg1
Transcriptasa inversa
3'
DONANTE1
5'
ADN1
F
3'
DBP1
CAS9-1
ARNsg1
RBP1
DONANTE1
5'
ADN1

FIG. 30

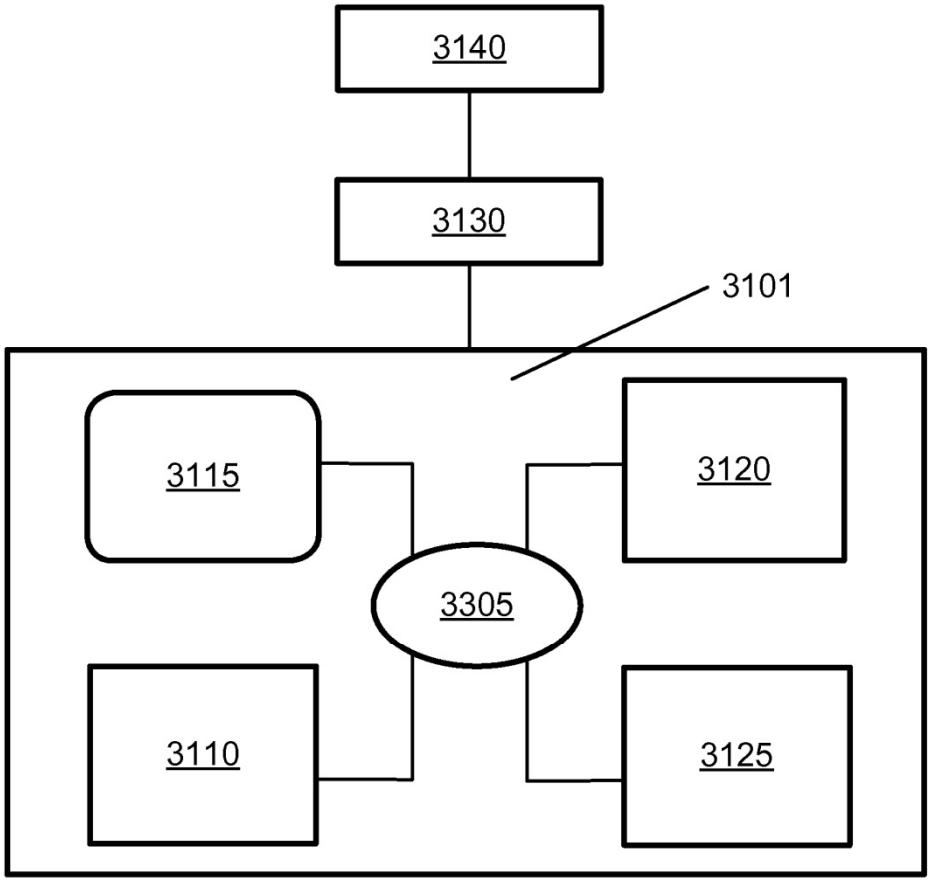


FIG. 31

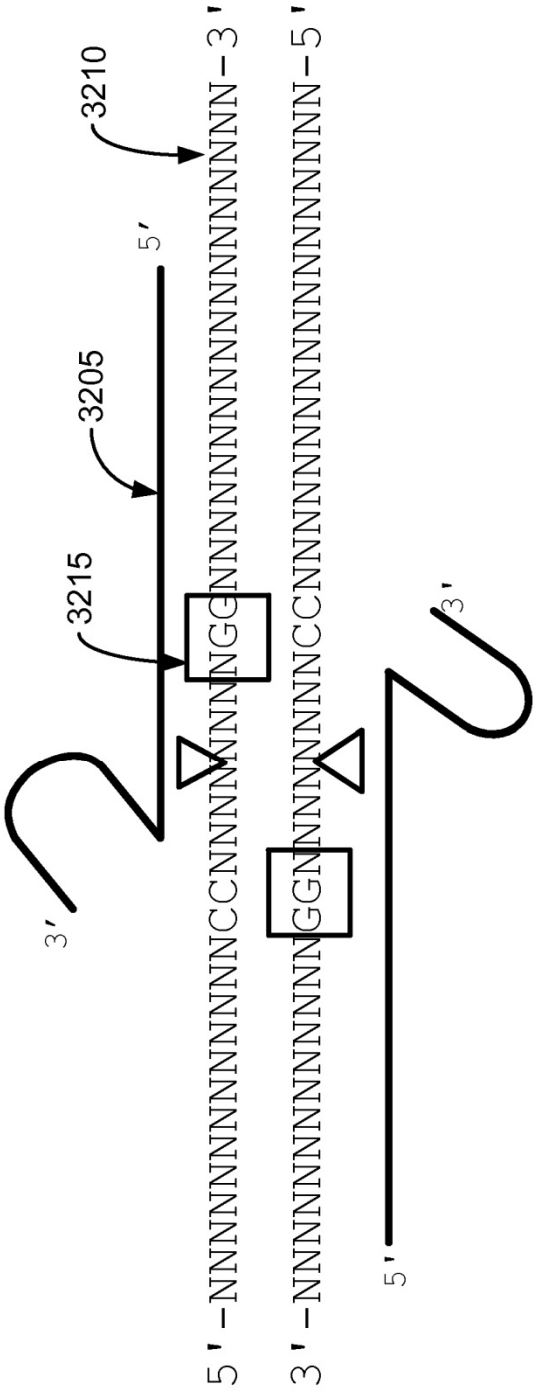

3210
3205
3215
5'
3'
5'-NNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
3'
5'

FIG. 32

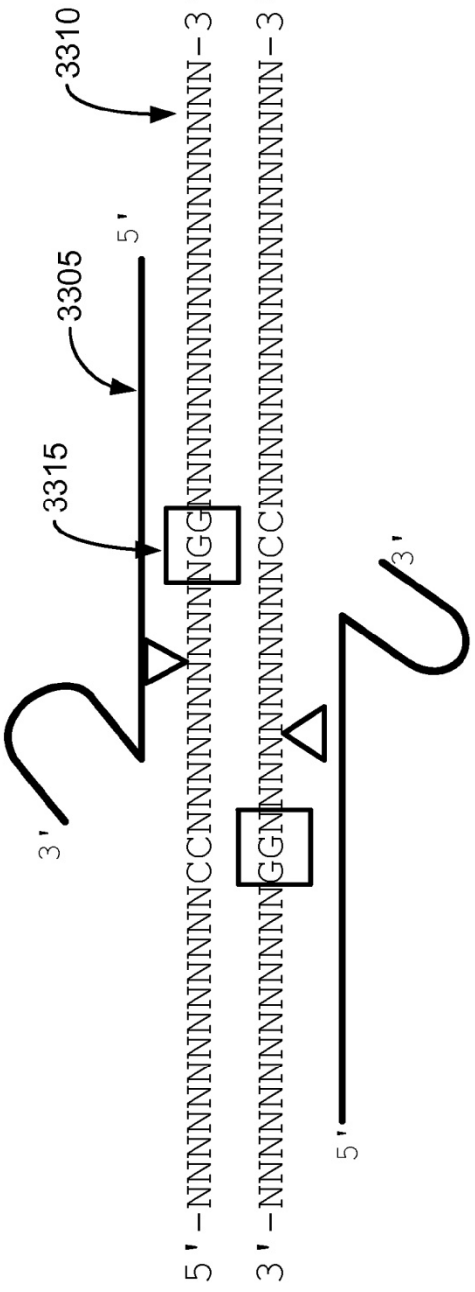

3310
3305
3315
5'
3'
5'-NNNNNNNNNNNNNNCCNNNNNNNNNNNNGGNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNNGGNNNNNNNNNNNNCCNNNNNNNNNNNNNNN-3'
5'
3'

FIG. 33

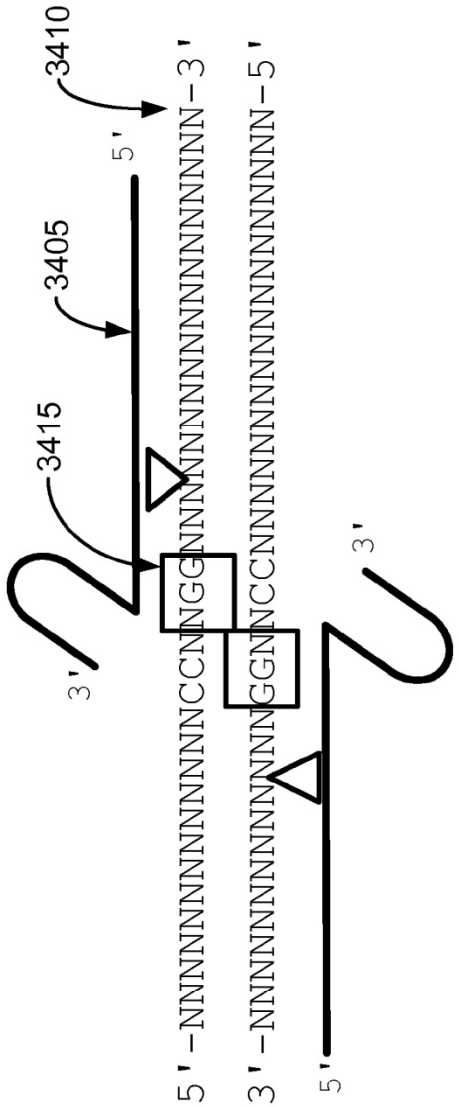

3410
3405
3415
5'
3'
5'-NNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN-5'
5'
3'

FIG. 34

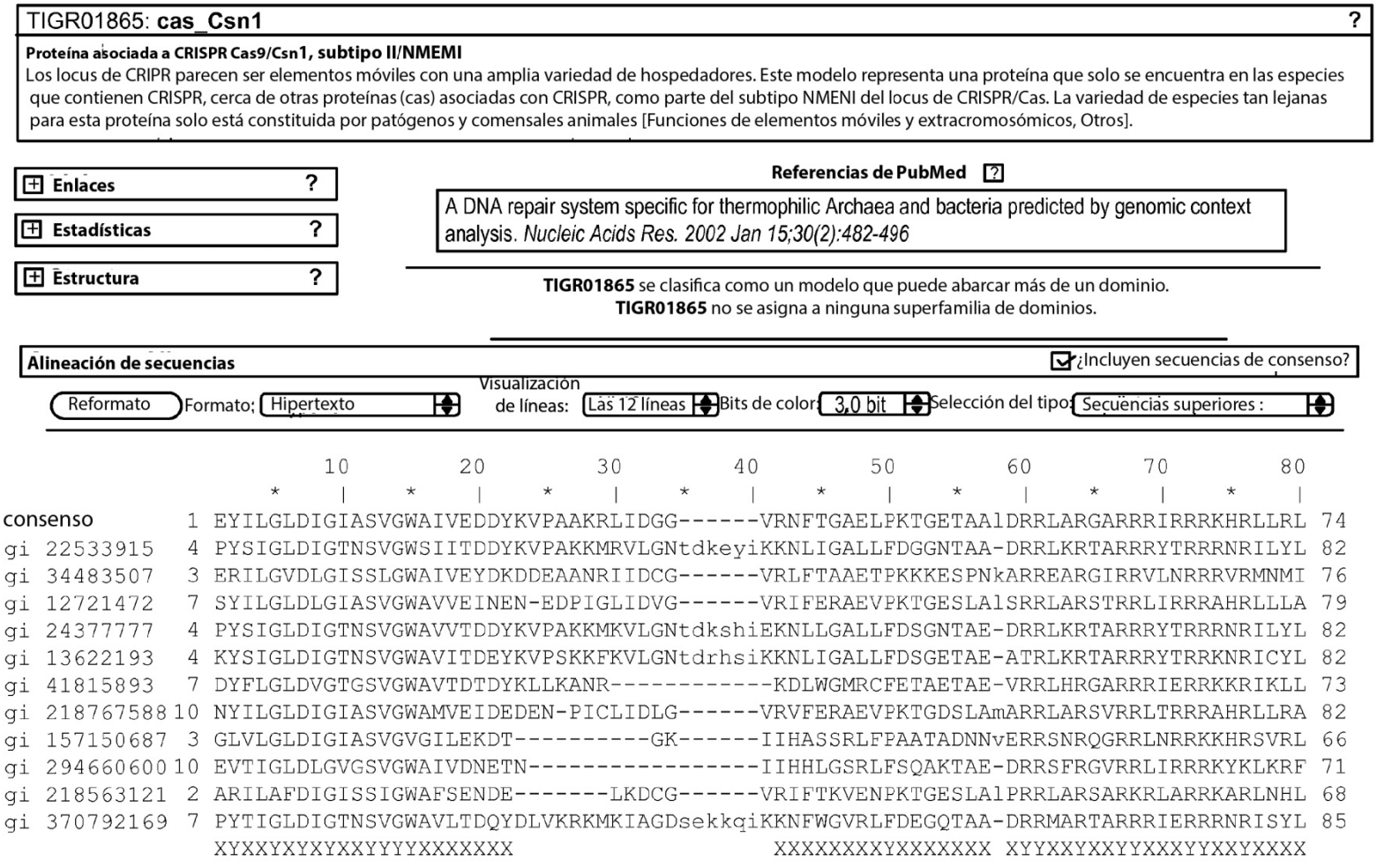

TIGR01865: cas_Csn1
?
Proteína asociada a CRISPR Cas9/Csn1, subtipo II/NMEMI
Los locus de CRIPR parecen ser elementos móviles con una amplia variedad de hospedadores. Este modelo representa una proteína que solo se encuentra en las especies que contienen CRISPR, cerca de otras proteínas (cas) asociadas con CRISPR, como parte del subtipo NMENI del locus de CRISPR/Cas. La variedad de especies tan lejanas para esta proteína solo está constituida por patógenos y comensales animales [Funciones de elementos móviles y extracromosómicos, Otros].
Enlaces ?
Estadísticas ?
Estructura ?
Referencias de PubMed ?
A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res. 2002 Jan 15;30(2):482-496
TIGR01865 se clasifica como un modelo que puede abarcar más de un dominio.
TIGR01865 no se asigna a ninguna superfamilia de dominios.
Alineación de secuencias
¿Incluyen secuencias de consenso?
Reformato
Formato: Hipertexto
Visualización de líneas: Las 12 líneas
Bits de color: 3,0 bit
Selección del tipo: Secuencias superiores :
10 20 30 40 50 60 70 80
consenso 1 EYILGLDIGIASVGWAIVEDDYKVPAAKRLIDGG------VRNFTGAELPKTGETAAlDRRLARGARRRIRRRKHRLLRL 74
gi 22533915 4 PYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNtdkeyiKKNLIGALLFDGGNTAA-DRRLKRTARRRYTRRRNRILYL 82
gi 34483507 3 ERILGVDLGISSLGWAIVEYDKDDEAANRIIDCG------VRLFTAAETPKKKESPNkARREARGIRRVLNRRRVRMNMI 76
gi 12721472 7 SYILGLDLGIASVGWAVVEINEN-EDPIGLIDVG------VRIFERAEVPKTGESLAlSRRLARSTRRLIRRRAHRLLLA 79
gi 24377777 4 PYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNtdkshiEKNLLGALLFDSGNTAE-DRRLKRTARRRYTRRRNRILYL 82
gi 13622193 4 KYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNtdrhsiKKNLIGALLFDSGETAE-ATRLKRTARRRYTRRKNRICYL 82
gi 41815893 7 DYFLGLDVGTGSVGWAVTDTDYKLLKANR-----------KDLWGMRCFETAETAE-VRRLHRGARRRIERRKKRIKLL 73
gi 218767588 10 NYILGLDIGIASVGWAMVEIDEDEN-PICLIDLG------VRVFERAEVPKTGDSLAmARRLARSVRRLTRRRAHRLLRA 82
gi 157150687 3 GLVLGLDIGIASVGVGILEKDT----------GK------IIHASSRLFPAATADNNvERRSNRQGRRLNRRKKHRSVRL 66
gi 294660600 10 EVTIGLDLGVGSVGWAIVDNETN----------------IIHHLGSRLFSQAKTAE-DRRSFRGVRRLIRRRKYKLKRF 71
gi 218563121 2 ARILAFDIGISSIGWAFSENDE-------LKDCG------VRIFTKVENPKTGESLAlPRRLARSARKRLARRKARLNHL 68
gi 370792169 7 PYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDsekkqiKKNFWGVRLFDEGQTAA-DRRMARTARRRIERRRNRISYL 85
XYXXYXYXYXXYYYYXXXXXXX XXXXXXXXYXXXXXXX XYYXXYXXYYXXXYYXXYXXXX

FIG. 35

```
                    90        100       110       120       130       140       150       160
                  *   |     *   |     *   |     *   |     *   |     *   |     *   |     *   |
consenso        75 QELFSREGSLTDFDFFsrLENSFLVEEDKRN----------------------------TIYHLRKAALENKLKPDE--LYLALLH 130
gi 22533915     83 QEIFAEEMSKVDDSFFhrLEDSFLVEEDKRGskypifatlq--eekdyhekfsTIYHLRKELADKKEKADLrlIYIALAH 160
gi 34483507     77 KKLFLRAGLIQDVDLD--GEGGMFYSKANRA----------------------DVWELRHDGLYRLLKGDE--LARVLLH 130
gi 12721472     80 KRFLKREGILSTIDLEk-----gLPNQA--------------------------WELRVAGLERRLSAIE--WGAVLLH 125
gi 24377777     83 QEIFSEEMGKVDDSFFhrLEDSFLVTEDKRGerhpifgnle--eevkyhenfpTIYHLRQYLADNPEKVDLrlVYLALAH 160
gi 13622193     83 QEIFSNEMAKVDDSFFhrLEESFLVEEDKKHerhpifgniv--devayhekypTIYHLRKKLVDSTDKADLrlIYLALAH 160
gi 41815893     74 QELFSQEIAKTDEGFFqrMKESPFYAEDKTIlqentlfndkdfacktyhkaypTINHLIKAWIENKVKPDPrlLYLACHN 153
gi 218767588    83 RRLLKREGVLQAADFDenGLIKSLP-----N----------------------TPWQLRAAALDRKLTPLE--WSAVLLH 133
gi 157150687    67 QDLFEGYGLLTDFSKVsmNLN----------------------------------PYQLRVQGMENQLTNEE--LFVALKN 111
gi 294660600    72 VNLIWKYNSYFGFKNKedILNNYQEQQKLHN----------------------TVLNLKLEALNAKIDPKA--LSWILHD 127
gi 218563121    69 KHLIANEFKLNYEDYQsfDESLAKAYKGSLI----------------------SPYELRFRALNELLSKQD--FARVILH 124
gi 370792169    86 QGIFAEEMSKTDANFFcrLSDSFYVDNEKRNsrhpffatie--eeveyhknypTIYHLREELVNSSEKADLrlVYLALAH 163
                   XXXYXXYXXXXXXXXX                                  YYYYXXXXXXXXXXXX  XXXXXXY
                          170       180       190       200       210       220       230       240
                        *   |     *   |     *   |     *   |     *   |     *   |     *   |     *   |
consenso        131 IIKHRGHFLIEGNDfdtAN----------------------------------------------------------- 149
gi 22533915     161 IIKFRGHFLIEDDSfdvRNtdiskqyqdfleifnttfennd1lsqnvdveailtdkisksakkdrilaqypnqkstgifa 240
gi 34483507     131 IAKHRGYKFIGDDE---AD----------------------------------------------------------- 146
gi 12721472     126 LIKHRGYLSKRKNEsqtNN----------------------------------------------------------- 144
gi 24377777     161 IIKFRGHFLIEG-KfdtRNndvqrlfqeflavydntfensslqeqnvqveeiltdkisksakkdrvlklfpneksngrfa 239
gi 13622193     161 MIKFRGHFLIEG-DlnpDNsdvdklfiqlvqtynqlfeenpinasgvdakailsarlsksrrlenliaqlpgekknglfg 239
gi 41815893     154 IIKKRGHFLFEG-DfdsENqfdtsiqalfeylredmevdidadsqkvkeilkdsslknsekqsrlnkilglkpsdkqkka 232
gi 218767588    134 LIKHRGYLSQRKNEgetAD----------------------------------------------------------- 152
gi 157150687    112 IVKRRGISYILDDASedgGT----------------------------------------------------------- 130
gi 294660600    128 YLKNRGYFYEDNRDfnvYP----------------------------------------------------------- 146
gi 218563121    125 IAKRRGYDDIKNSD----D----------------------------------------------------------- 139
gi 370792169    164 IIKYRGNFLIEG-AldtQNtsvdgiykqfiqtynqvfasgiedgelkkledn kdvakilvekvtrkeklerilklypgek 242
                    XXYXYYYXXXXXX X     X
```

FIG. 35
(Continuación)

```
                    250       260       270       280       290       300       310       320
                  *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |
consenso      150 ----------------------------------------------------------------------KET 152
gi 22533915   241 eflkl------ivgnqadfkkyfnledktplqfakdsydedlenllgqigdefadlfsaakklydsvllsgiltvidLST 314
gi 34483507   147 ----------------------------------------------------------------------EES 149
gi 12721472   145 ----------------------------------------------------------------------KEL 147
gi 24377777   240 eflkl------ivgnqadfkkhfeleekaplqfskdtyeeelevllaqigdnyaelflsakklydsillsgiltvtdVGT 313
gi 13622193   240 nlial------slgltpnfksnfdlaedaklqlskdtyddddldnllaqigdqyadlflaaknlsdaillsdilrvntEIT 313
gi 41815893   233 itnl-------isgnkinfadlydnpdlkdaeknsisfskddfdalsddlasilgdsfelllkakavyncsvlskvIGD 304
gi 218767588  153 ----------------------------------------------------------------------KEL 155
gi 157150687  131 ----------------------------------------------------------------------VS 132
gi 294660600  147 ----------------------------------------------------------------------T 147
gi 218563121  140 ----------------------------------------------------------------------KEK 142
gi 370792169  243 sagmfaqfislivgskgnfqkpfdlieksdiecakdsyeedlesllaligdeyaelfvaaknaysavvlssiitvaeTET 322

                    330       340       350       360       370       380       390       400
                  *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |
consenso      153 GALLSAVMINRYLEHEADLRTLKELILKKFPKKYKEifse---------------------------------- 192
gi 22533915   315 KAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEifadsskdgyagyyiegktnqeafyk---------------- 375
gi 34483507   150 GKVKKAGVVLRQNFEAAGCRTVGEWLWRERGANGKK---------------------------------- 185
gi 12721472   148 GALLSGVAQNHQLLQSDDYRTPAELALKKFAKEEGHi---------------------------------- 184
gi 24377777   314 KAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEvfsdvskdgyagyyidgktnqeafyk---------------- 374
gi 13622193   314 KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEiffdqskngyagyyidggasqeefyk---------------- 374
gi 41815893   305 EQYLSFAKVKIYEKHKTDLTKLKNVIKKHFPKDYKKvfgynkneknnnnysgyvgvcktkskkliinnsvnqedfykflk 384
gi 218767588  156 GALLKGVADNAHALQTGDFRTPAELALNKFEKESGHi---------------------------------- 192
gi 157150687  133 SDYGKAVEENRKLLAEKTPGQIQLERFEKYGQLRGDftv---------------------------------- 171
gi 294660600  148 EELANYFDEFGYYKGIIDSKNDDDDKLEEGLTKYKFsnqh---------------------------------- 187
gi 218563121  143 GAILKA-IKQNEEKLANYQSVGEYLYKEYFQKFKEnske---------------------------------- 180
gi 370792169  323 KANLSASMIERFDTHEEDLGELKAFIKLHLPKHYEEifsntekhgyagyyidgktkqadfyk---------------- 383
                  XXXXXX  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

FIG. 35
(Continuación)

```
                                410       420       430       440       450       460       470       480
                              *   |     *   |     *   |     *   |     *   |     *   |     *   |     *   |
consenso                193 ------------------------tflRNQRGFYNGSIPRQLLLEELEAIFRKQReyypf----------------ikllTFR 235
gi 22533915             376 ylskl1tkqedsenflekiknedflRKQRTFDNGSIPHQVHLTELKAIIRRQSeyypflken----------qdriekilTFR 448
gi 34483507             186 ------------------------RNKHGDYEISIHRDLLVEEVEAIFVAQQ-----------------------EMR 216
gi 12721472             185 ------------------------RNQRGAYTHRFNRLDLLAELNLLFAQQH-----------------------QFG 215
gi 24377777             375 ylkgllnkiegsgyfldkieredflRKQRTFDNGSIPHQIHLQEMRAIIRRQAefypfladn----------qdriekllTFR 447
gi 13622193             375 fikpilekmdgteellvklnredllRKQRTFDNGSIPHQIHLGELHAILRRQEdfypflkdn----------rekiekilTFR 447
gi 41815893             385 tilsakseikevndilteietgtflPKQISKSNAEIPYQLRKMELEKILSNAEkhfsflkqkdekglshsekiimllTFK 464
gi 218767588            193 ------------------------RNQRGDYSHTFSRKDLQAELILLFEKQK-----------------------EFG 223
gi 157150687            172 ------------------------eeNGEKHRLINVFSTSAYRKEAERILRKQQefnskitde----------fiedyliilTGK 222
gi 294660600            188 ------------------------wleEVKKVLSNQTGLPEKFKEEYESLFSYVRnyseg----------pgsiNSV 230
gi 218563121            181 ------------------------ftnvRNKKESYERCIAQSFLKDELKLIFKKQR-----------------------EFG 215
gi 370792169            384 ymkmtleniegadyfiakiekenflRKQRTFDNGAIPHQLHLEELEAILHQQAkyypflken----------ydkikslvTFR 456
                                                                YXYXXXXXXXXYXXXXXXXYXXXYXXXYX          XYX

                                490       500       510       520       530       540       550       560
                              *   |     *   |     *   |     *   |     *   |     *   |     *   |     *   |
consenso                236 IPYYIGPLAEGKSEFA----------------------------------FVDKPASAenfIENMTGKCTYLPEEKRAPKHSLLAEKFTV 291
gi 22533915             449 IPYYIGPLAREKSDFAwmtrktdds----------irpwnfedLVDKEKSAeakIHRMTNNDFYLPEEKVLPKHSLIYEKFTV 521
gi 34483507             217 STIATDALKAAYREIA----------------------------FFVRPMQR---IEKMVGHCTYFPEERRAPKSAPTAEKFIA 269
gi 12721472             216 NPHCKEHIQQYMTELL----------------------------MWQKPALSgeaILKMLGKCTHEKNEFKAAKHTYSAERFVW 271
gi 24377777             448 IPYYVGPLARGKSDRAwlsrksadk----------itpwnfdeIVDKESSAeakINRMTNYDLYLPNQKVLPKHSLLYEKFTV 520
gi 13622193             448 IPYYVGPLARGNSRFAwmtrkseet----------itpwnfeeVVDKGASAqskIERMTNFDKNLPNEKVLPKHSLLYEYFTV 520
gi 41815893             465 IPYYIGPINDNHKKFFpdrcwvvkkekspsgkttpwnffdHIDKEKTAeafITSRTNFCTYLVGESVLPKSSLLYSEYTV 544
gi 218767588            224 NPHVSGGLKEGIETLL----------------------------MTQRPALSgdaVQKMLGHCTFEPAEPKAAKNTYTAERFIW 279
gi 157150687            223 RKYYHGPGNEKSRTDYg----------------------------rFRTDGTTLdniFGILIGKCTFYTEEYRASKASYTAQEFNL 280
gi 294660600            231 SPYGIYHLDEKEGKVV----------------------------QKYnniWDKTIGKCSIFPDEYRAPKNSPIAMIFNE 281
gi 218563121            216 FSFS-KKFEEEVLSVA----------------------------FYKRALKD---FSHLVGNCSFFTDEKRAPKNSPLAFMFVA 267
gi 370792169            457 IPYFVGPLANGQSEFAwltrkadge----------irpwnieeKVDFGKSAvdfIEKMTNKDTYLPKENVLPKHSLCYQKYLV 529
                                XXYX XXXXXXXXXXX                                XXX   YXXXYXYXYXXXXXXYXXXYYYXXXYXXYXX
```

FIG. 35
(Continuación)

```
                                570       580       590       600       610       620       630       640
                              *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |
consenso                  292 LNELNNVR---IIILeQGETKILSKEEKQELLDLLFKKKKLTYKKLRKLLGLSEDAIFKGLRYEG---------- 353
gi 22533915               522 YNELTKVR---YKNE-QGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIG---------- 582
gi 34483507               270 ISKFFSTV---IIDNeGWEQKIIERKTLEELLDFAVSREKVEFRHLRKFLDLSDNEIFKGLHYKGkpktakkreatlfdp 346
gi 12721472               272 LTKLNN-----LRILeDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKNLRYS---------- 330
gi 24377777               521 YNELTKVK---YKTE-QGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTG---------- 581
gi 13622193               521 YNELTKVK---YVTEgMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG---------- 582
gi 41815893               545 LNEINNLQ---IIID---GKNICDIKLKQKIYEDLFKKYKKITQKQISTFIKHEGICNKTDEVII---------- 603
gi 218767588              280 LTKLNN-----LRILeQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRY-G---------- 338
gi 157150687              281 ----LNDL---NNLTvPTETKKLSEEQKKLIIEYAKSAKTLGASTLLKYIAKMIDASVDQIRGYR---------- 338
gi 294660600              282 INELSTIRsysIYLTgWFINQEFKKAYLNKLLDLLIKTNSEKPIDARQFKKLREETIAESIGKETlkdveseeklek-dd 360
gi 218563121              268 LTRIINL-----LNNLkNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKG---------- 323
gi 370792169              530 YNELTKVR---YIND-QGKTSYFSGQEKEQIFNDLFKQKRKVK-KKDLELFLRNMSHVESPTIEG---------- 589
                                  XX        XXXX    XXXXXXXXXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXX

                                650       660       670       680       690       700       710       720
                              *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |   *   |
consenso                  354 ----------LDNAEKAFNISLKTYHLKRKALgDKDLLLDn----PKNPKDLDEIVKILTLYKDREm----------iK 407
gi 22533915               583 ----------LDKENKAFNASLGTYHDLEKIL-DKDFLDn----PDNESILEDIVQTLTLFEDREm----------iK 635
gi 34483507               347 neptelefdKVEAEKKAWISLRGAAKLREALgNEFYGRf----VALGKHADEATKILTYYKDEG----------Q 407
gi 12721472               331 ----------KENAESATFMELKAWHAIRKALeNQGLKDtwqdlAKKPDLLDEIGTAFSLYKTDE----------D 386
gi 24377777               582 ----------LDKENKVFNASYGTYHDLCKIL-DKDFLDn----SKNEKILEDIVLTLTLFEDREm----------iR 634
gi 13622193               583 ----------VEDR---FNASLGTYHDLLKIIkDKDFLDn----EENEDILEDIVLTLTLFEDREm----------iE 633
gi 41815893               604 ----------LGID-KECTSSLKSYIELKNIF--GKQVDe----ISTKNMLEEIIRWATIYDEGEgkt----------ilkT 658
gi 218767588              339 ----------KDNAEASTLMEMKAYHAISRALeKEGLKDkksplNLSPELQDEIGTAFSLFKTDE----------D 394
gi 157150687              339 ----------VDVNNKPEMHTFEVYRKMQSLEtIKVEELp--------RKVLDELAHILTLNTEREgi----------eE 390
gi 294660600              361 hkwklkglklNTNGKIQYNDLSSLAKFVHKLkQHLKLDflledQYTPLDKINFLQSLYVYLGKHlrysnrvdsanlkefS 440
gi 218563121              324 ----------EKGTYFIEFKKYKEFIKALgEHNL--------SQDDLNEIAKDITLIKDEi----------K 367
gi 370792169              590 ----------LEDS---FNSSYSTYHDLLKVGiKQEILDn----PVNTEMLENIVKILTVFEDKRm----------iK 640
                                            X   XXXXXXXXYYXXXXXXX  XXX       XXXXXXXXXXXXYXXXXX       X
```

FIG. 35
(Continuación)

```
                730       740       750       760       770       780       790       800
                *    |    *    |    *    |    *    |    *    |    *    |    *    |    *    |
consenso    408 KRLELYKDVLNEEQ--VKKLVRLH-------FTGWGRLSLKALRGIRPLMEQGKRYDEAILELGGNrnfmgNINDSQLLP 478
gi 22533915 636 KRLENYKDLFTESQ--LKKLYRRH-------YTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNrnfmqLINDDGLSF 706
gi 34483507 408 KRRELTKLPLEAEM--VERLVKIG-------FSDFLKLSLKAIRDILPAMESGARYDEAVLMLGVP-----HKEKSAILP 473
gi 12721472 387 IQQYLTNKVPNSVI--NALLVSLN-------FDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHhygeaNQKTSQLLP 457
gi 24377777 635 KRLENYSKLLTKEQ--VKKLERRH-------YTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNrnfmqLINDDALSF 705
gi 13622193 634 ERLKTYAHLFDDKV--MKQLKRRR-------YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANrnfmqLINDDSLTF 704
gi 41815893 659 KIKAEYGKYCSDEQ--IKKILNLK-------FSGWGRLSRKFLETVTSEMPGFSEPVNIITAMRETqnnlmELLSSEFTF 729
gi 218767588 395 ITGRLKDRIQPEIL--EALLKHIS-------FDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDhygkkNTEEKIYLP 465
gi 157150687 391 AINSKLKDIFNRDQ--VLELVQFRknnsslfSKGWHNFSIKLMMELIPELYETSEEQMTILTRLGKqrskeTSKRTKYID 468
gi 294660600 441 DSSRLFERVLQEQKdgLFKLFEQTdkddekiLTQTHSLSTKAMLLAITRMTNLDNDEDNQKNNDKGwnfeaIKNFDQKFI 520
gi 218563121 368 LKKALAKYDLNQNQ--IDSLSKLE-------FKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKva--iNEDKKDFLP 436
gi 370792169 641 EQLQQFSDVLDFVV--LKKLERRH-------YTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGINrnlmqLINDSNLSF 711
                XXXXXXXXXXXXXX  XXXYXXXX       XXXYXXXYXXXXXXXXXXXXXXXXXXXXXXXX      XXXXXXXXX
                          810       820       830       840       850       860       870       880
                          *    |    *    |    *    |    *    |    *    |    *    |    *    |    *    |
consenso    479 KINITk----------------aKDEILNPVVKRALLQARKVVNELVKKYG---PPDKIVIEMARe----E----QGTN 530
gi 22533915 707 KSIISkaqagshs----dnlkevvGELAGSPAIKKGILQSLKIVDELVKVMGy--EPEQIVVEMARe----N----QTTN 772
gi 34483507 474 PLNKT----------------DIDILNPTVIRAFAQFRKVANALVRKYG---AFDRVHIFELAR----E----INTK 522
gi 12721472 458 AIPAQ----------------EIRNPVVLRTLSQARKVINAIIRQYG---SPARVHIETGR----E----LGKS 504
gi 24377777 706 KEEIAkaqviget----dnlnqvvSDIAGSPAIKKGILQSLKIVDELVKIMGh--QPENIVVEMARe----N----QFTN 771
gi 13622193 705 KEDIQkaqvsgqg----dslhehiANLAGSPAIKKGILQTVKVVDELVKVMGr-hKPENIVIEMARe----N----QTTQ 771
gi 41815893 730 TENIKkinsgfedaekqfsydglvKPLFLSPSVKKMLWQTLKLVKEISHITQa--PPKKIFIEMAXgaeleP----ARTK 803
gi 218767588 466 PIPA----------------DEIRNPVVLRALSQARKVINGVVRRYG---SPARIHIETARe----------VGKS 512
gi 157150687 469 EKELT----------------EEIYNPVVAKSVRQAIKIINEATKKYG---IFDNIVIEMARe----NneedAKKD 521
gi 294660600 521 DITKTnnnlslkqd-krylddqfiNDAILSPGVKRILREATKVFNAILKQFSeeyDVTKVVIELARel--sE----EKEL 593
gi 218563121 437 AFNETy----------------yKDEVTNPVVLRAIKEYRKVLNALLKKYG---KVHKINIELAR----E----VGKN 487
gi 370792169 712 KSIIEkeqvttad----kdiqsivADLAGSPAIKKGILQSLKIVDELVSVMGy--PPQRIVVEMARe----N----QTTG 777
                XXXX                XXXYYXYXYXXXYXXYXXXXXYY   XYXXYXYYYYYY      XXXX
```

FIG. 35
(Continuación)

```
                                890       900       910       920       930       940       950       960
                                 *    |    *    |    *    |    *    |    *    |    *    |    *    |
consenso                   531 FGKRNSKERYKKNEDKIKEFASALGKEILKEEPTensSKNILKLRLYYQQNGKCMYTGKEIDIDdl-fDLSYYEIDHILP 609
gi 22533915                773 QGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTdnqALQNERLFLYYLQNGRDMYTGEALDID---NLSQYDIDHIIP 848
gi 34483507                523 GEIEDIKESQRKNEKERKEAADWIAETSFQVPLT---RKNILKKRLYIQQDGRCAYTGDVIELErl-fDEGYCEIDHILP 598
gi 12721472                505 FKERREIQKQQEDNRTKRESAVQKFKELFSDFSSepkSKDILKFRLYEQQHGKCLYSGKEINIHrl-nEKGYVEIDHALP 583
gi 24377777                772 QGRRNSQQRLKGLTDSIKEFGS----QILKEHPVensQLQNDRLFLYYLQNGRDMYTGEELDID----YLSQYDIDHIIP 843
gi 13622193                772 KGQKNSRERMKRIEEGIKELGS----QILKEHPVentQLQNEKLYLYYLQNGRDMYVDQELDIN----RLSDYDVDHIVP 843
gi 41815893                804 TRLKILQDLYNNCKNDADAFSSEIKDLSGKIENEdnlRLRSDKLYLYYTQLGKCMYCGKPIEIGhv-fDTSNYDIDHIYP 882
gi 218767588               513 FKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGepkSKDILKLRLYEQQHGKCLYSGKEINLGrl-nEKGYVEIDHALP 591
gi 157150687               522 YIKRQKANQDEKNAAMEKAAFQYNGKKELPDNIFhghKELTTKIRLWHQQGEKCLYTGKNIPISdlihNQYKYEIDHILP 601
gi 294660600               594 ENTKNYKKLIKKNGDKISEGLKALGIAEDKIEEIlkePTKSYKVLLWLQQDHIDPYSQKEIAFDdiltKTEKTEIDHIIP 673
gi 218563121               488 -HSQRAKIEKEQNENYKAKKDAELECEKLGLKIN---SKNILKLRLFKEQKEFCAYSGEKIKISdl-qDEKMLEIDHIYP 562
gi 370792169               778 KGKNNSRPRYKSLEKAIKEFGS----QILKEHPTdnqELRNNRLYLYYLQNGKDMYTGQDLDIH----NLSNYDIDHIVP 849
                                 XXXXXXXXXXXXXXXXXXXXXX    XXXXXXXX  XXXXXYXXYYXXXYXXYXXYXYX    XXXXXYYYYYXY
                                      970       980       990      1000      1010      1020      1030      1040
                                       *    |    *    |    *    |    *    |    *    |    *    |    *    |
consenso                   610 QSRSFDDSISNKVLVLASENQEKGDQTPYEaEIVkkDSAFWNKFEAYVLISKR----------KSDKLT----RAERGGL 675
gi 22533915                848 QAFIKDDSIDNRVLVSSAKNRGKSDDVPSL-EIVkdCKVFWKKLLDAKLMSQR----------KYDNLT----KAERGGL 913
gi 34483507                599 RSRSADDSFANKVLCLARANQQKTDRTPYEwFGH--DAARWNAFETRTSAPSN----------RVRTCKgkidRLLKKNF 666
gi 12721472                584 FSRTWDDSFNNKVLVLASENQNKGNQTPYEwLQGkiNSERWKNFVALVLGSQC----------SAAKKQ----RLLTQVI 649
gi 24377777                844 QAFIKDNAIDNRVLTSSKENRGKSDDVPSK-DVVrkMKSYWSKLLSAKLITQR----------KFDNLT----KAERGGL 908
gi 13622193                844 QSFLKDDSIDNKVLTRSDKNRGKSDNVPSE-EVVkkMKNYWRQLLNAKLITQR----------KFDNLT----KAERGGL 908
gi 41815893                883 QSKIKDDSISNRVLVCSSCNKNKEDKYPLKsEIQskQRGFWNFLQRNNFISLE----------KLNRLT----RATP--I 946
gi 218767588               592 FSRTWDDSFNNKVLVLGSENQNKGNQTPYEyFNGkdNSREWQEFKARVETSRF----------PRSKKQ----RILLQKF 657
gi 157150687               602 LSLSFDDSLSNKVLVLATANQEKGQRTPFQaLDSmdDAWSYREFKSYVKDSKL----------LSNKKKdy-lLTEEDIS 670
gi 294660600               674 YSISFDDSSSNKLLVLAESNQAKSNQTPYEfINSgkAEITWEVYEAYCHKFKNgdssslldstqRSKKFA----KMMKTDT 749
gi 218563121               563 YSRSFDDSYMNKVLVFTKQNQEKLNQTPFEaFGN--DSAKWQKIEVLAKNLPT----------KKQK-----RILDKNY 624
gi 370792169               850 QSFITDNSIDNLVLTSSAGNREKGDDVPPL-EIVrkRKVFWEKLYQGNLMSKR----------KFDYLT----KAERGGL 914
                                XYXXXXYYYXXYYXXXXXXYYXYXXYXXX XXX  XXXXYXXXXXXXXXXXXXX    XXXX      XXXX  X
```

FIG. 35
(Continuación)

```
                1050      1060      1070      1080      1090      1100      1110      1120
                  *    |    *    |    *    |    *    |    *    |    *    |    *    |
consenso        676 SDDDKAGFIDRNLNDTRYITRVVANYLKDRFNFHl----------KKRKVKVVTLKGQLTSQLRKKWG----LYKKR 738
gi 22533915     914 TSDDKARFIQRQLVETRQITKHVARILDERFNNEld----------kgrRIRKVKIVTLKSNLVSNFRKEFG----FYKIR 981
gi 34483507     667 DENSEMAFKDRNLNDTRYMARAIKTYCEQYWVFKn----------SHTKAPVQVRSGKLTSVLRYQWG----LESKD 729
gi 12721472     650 DDN---KFIDRNLNDTRYIARFLSHYIQENLLLVg----------KNKKNVFTPNGQITALLRSRWG----LIKAR 708
gi 24377777     909 TDDDKAGFIKRQLVETRQITKHVARILDERFNTEtde----------nnkKIRQVKIVTLKSNLVSNFRKEFE----LYKVR 976
gi 13622193     909 SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKyde----------ndkLIREVKVITLKSKLVSDFRKDFQ----FYKVR 976
gi 41815893     947 SDDETAKFIARQLVETRQATKVAAKVLEKMFP----------ETKIVYSKAETVSMFRNKFD----IVKCR 1003
gi 218767588    658 DED---GFKERNLNDTRYVNRFLCQFVADRMRLTg----------KGKKRVFASNGQITNLLRGFWG----LRKVR 716
gi 157150687    671 KIEVKQKFIERNLVDTRYSSRVVLNALQDFYKSHq----------LDTTISVVRGQFTSQLRRKWG----IEKSR 731
gi 294660600    750 SSKYDIGFLARNLNDTRYATIVFRDALKDYANNHlv----------edKPMFKVVCINGGVTSFLRKNFDpkswYAKKD 818
gi 218563121    625 KDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLplsddentklndtqkGSKVHVEAKSGMLTSALRHTWG----FSAKD 700
gi 370792169    915 TEADKARFIHRQLVETRQITKNVANILHQRFNYEkdd----------hgnTMKQVRIVTLKSALVSQFRKQFQ----LYKVR 982
                    XXX   XYXXYYYYYYYYXXXXXXXXXXXXXXX          XXXYXXXXXXXXYYXYYXXYX    XXYXY

                1130      1140      1150      1160      1170      1180      1190
                  *    |    *    |    *    |    *    |    *    |    *    |
consenso        739 EINNYHHAHDAYINAVSTNALVKKFSQLEpefRYKEYHNFDGRKkkksa----------tdkkvKFSNPMEFFKQKV 805
gi 22533915     982 EVNNYHHAHDAYLNAVVAKAILTKYPQLEpefVYGDYPKYNSYKtrksa----------teklfFYSNIMNFFKTKV 1048
gi 34483507     730 RESHTHHAVDAIIIAFSTQGMVQKLSEYY---RFKETHREKERP----------KLAVPLANFRDAV 783
gi 12721472     709 ENNNRHHALDAIVVACATPSMQQKITRFI---RFKEVHPYKIENryemvdqe--sgeiispHFPEPWAYFRQEV 777
gi 24377777     977 EINDYHHAHDAYLNAVIGKALLGVYPQLEpefVYGDYPHFHGHIenkat----------akkfFYSNIMNFFKKDD 1042
gi 13622193     977 EINNYHHAHDAYLNAVVGTALIKKYPKLEsefVYGDYKVYDVRKmiakseqeigkatakyfFYSNIMNFFKTEI 1050
gi 41815893    1004 EINDFHHAHDAYLNIVVGNVYNTKFTNNPwnfIKEKRDNPKIADtynyyk----vfdydvKRNNITAWEKGKT 1072
gi 218767588    717 AENDRHHALDAVVVACSTVAMQQKITRFV---RYKEMNAFDGKTidketge----vlhqktHFPQPWEFFAQEV 783
gi 157150687    732 ETYHHHAVDALIIAASSQLRLWKKHSNPLiayKEGQFVDSETGEi----------vSLSDEEYKELV 788
gi 294660600    819 RDKNIHHAVDASIISIFSNETKTLFNQLTkfaDYKLFKNTDGSWkkidp----------ktgvVSEVTDENWKQIr 884
gi 218563121    701 RNNHLHHAIDAVIIAYANNSIVKAFSDFK---KEQESNSAELYAkkisel----------dyknkrKFFEPFSGFRQKV 766
gi 370792169    983 DVNDYHHAHDAYLNGVVANTLLKVYPQLEpefVYGDYHQFDWFKankat----------akkqFYTNIMLFFAQKD 1048
                    XXYYXYYYYXYYXXXXXXXXXXXXXXXXX   XYXXXXXXXXXXX                XXXXXXYXXX
```

FIG. 35
(Continuación)

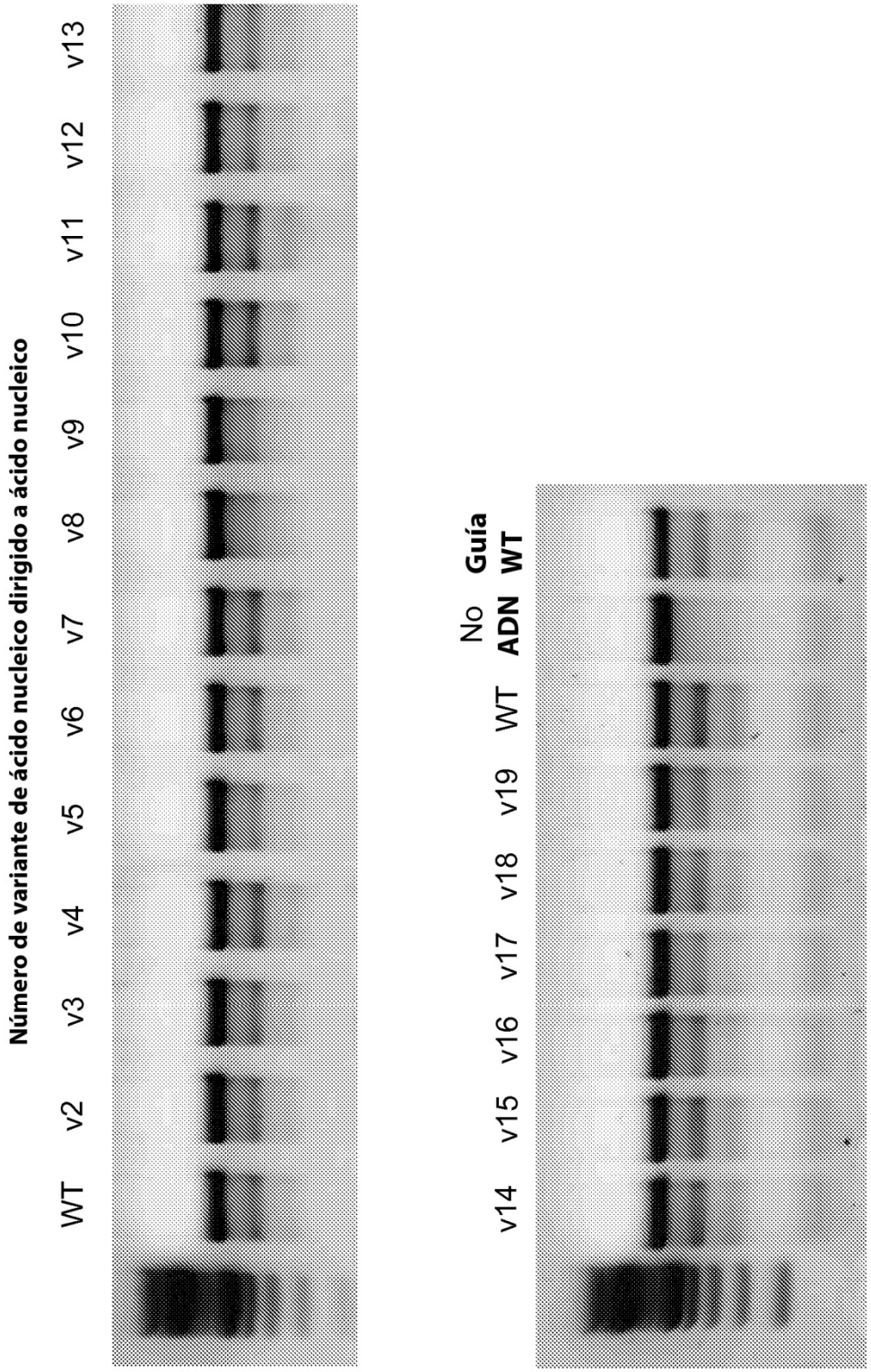

Número de variante de ácido nucleico dirigido a ácido nucleico
WT
v2
v3
v4
v5
v6
v7
v8
v9
v10
v11
v12
v13
v14
v15
v16
v17
v18
v19
WT
No ADN
Guía WT

FIG. 36

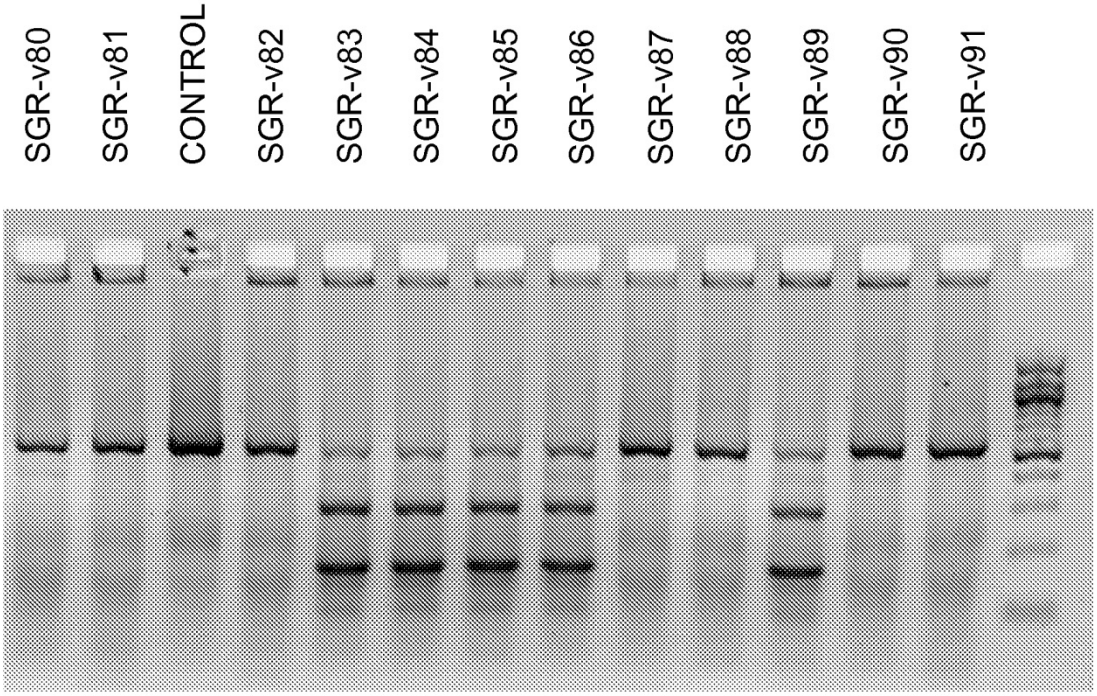

SGR-v80
SGR-v81
CONTROL
SGR-v82
SGR-v83
SGR-v84
SGR-v85
SGR-v86
SGR-v87
SGR-v88
SGR-v89
SGR-v90
SGR-v91

FIG. 37A

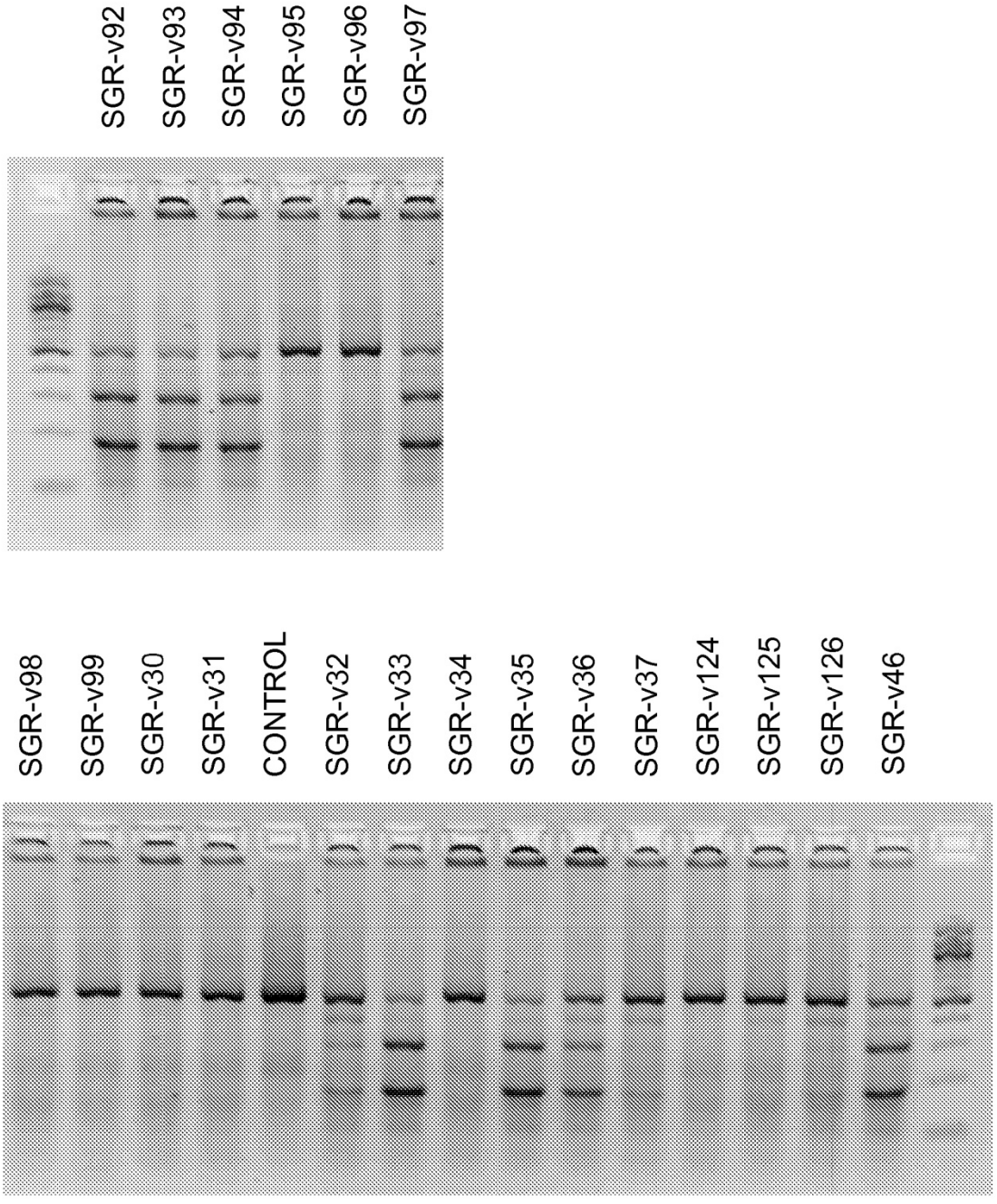

SGR-v92
SGR-v93
SGR-v94
SGR-v95
SGR-v96
SGR-v97
SGR-v98
SGR-v99
SGR-v30
SGR-v31
CONTROL
SGR-v32
SGR-v33
SGR-v34
SGR-v35
SGR-v36
SGR-v37
SGR-v124
SGR-v125
SGR-v126
SGR-v46

FIG. 37B

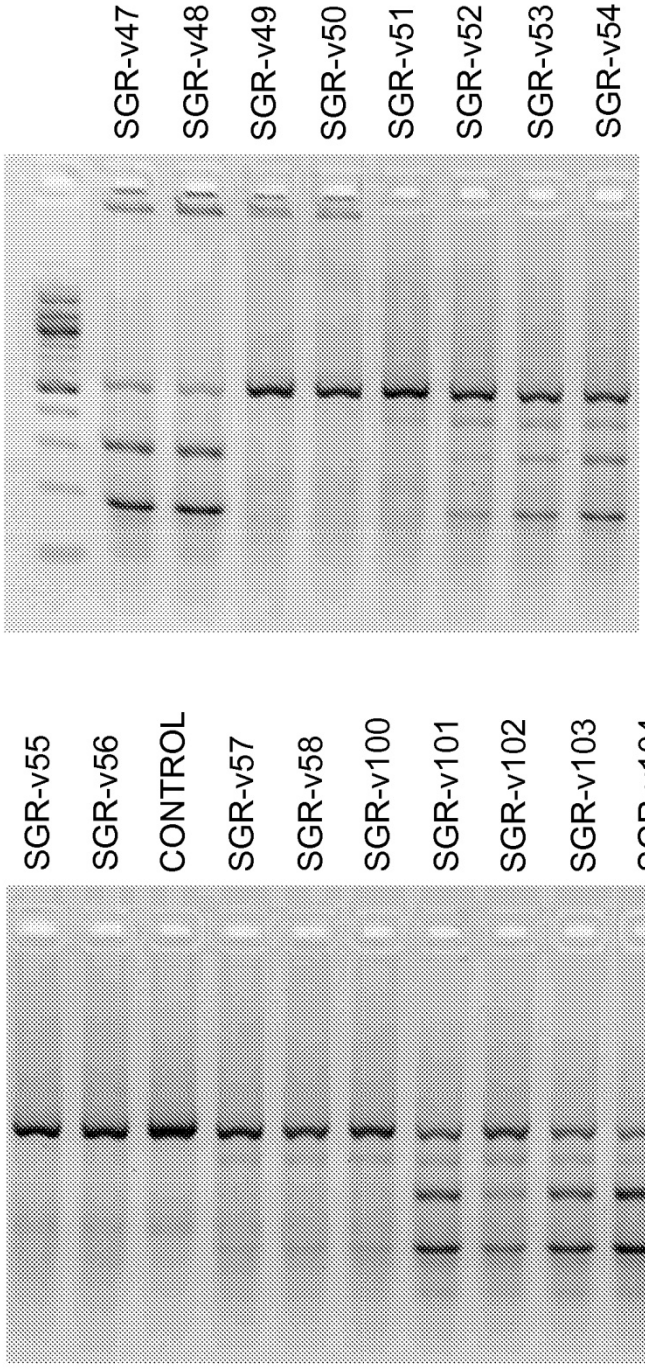

SGR-v47
SGR-v48
SGR-v49
SGR-v50
SGR-v51
SGR-v52
SGR-v53
SGR-v54
SGR-v55
SGR-v56
CONTROL
SGR-v57
SGR-v58
SGR-v100
SGR-v101
SGR-v102
SGR-v103
SGR-v104
SGR-v105
SGR-v106
SGR-v107

FIG. 37C

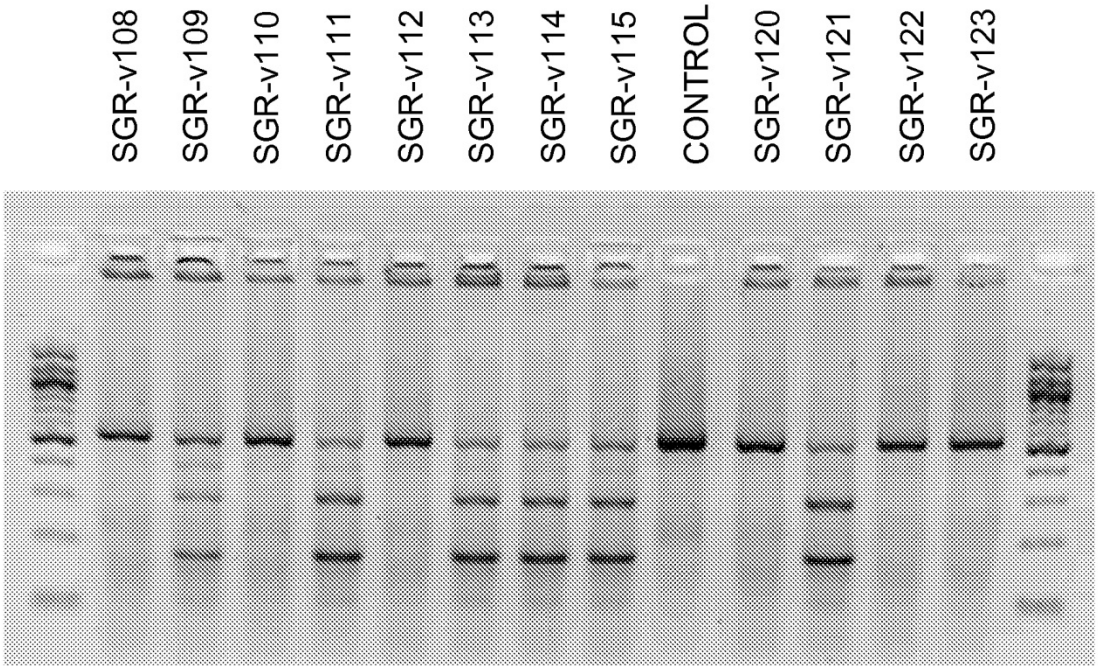

SGR-v108
SGR-v109
SGR-v110
SGR-v111
SGR-v112
SGR-v113
SGR-v114
SGR-v115
CONTROL
SGR-v120
SGR-v121
SGR-v122
SGR-v123

FIG. 37D

Secuencia de Csy4

>gi|107101871|ref|ZP 01365789.1| proteína hipotética PaerPA 01002916 [Pseudomonas aeruginosa PACS2]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLRALLARPWLEGLRDHLQFGEPAVVPHPTPYRQVS
RVQVKSNPERLRRRLMRRHELSEEEARKRIPDTVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF

>gi|254235433|ref|ZP 04928756.1| proteína hipotética PACG 01340 [Pseudomonas aeruginosa C3719]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLRAL
LARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTVARTLD
LPFVTLRSQSTGQHFRLFIRHGPLQATAEEGGFTCYGLSKGGFVPWF

>gi|254240857|ref|ZP 04934179.1| proteína hipotética PA2G 01531 [Pseudomonas aeruginosa 2192]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLHAL
LARPWLEGLRDHLQFGEAAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTVARTLD
LPFVTLRSQSTGQHFRLFIRHGPLQATAEEGGFTCYGLSKGGFVPWF

FIG. 38

His29Ala

```
   1 mdhyldirlr pdpefppaql msvlfgkl**a**q alvaqggdri gvsfpdldes rsrlgerlri
  61 hasaddlral larpwleglr dhlqfgepav vphptpyrqv srvqvksnpe rlrrrlmrrh
 121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfrlfir hgplqataee ggftcyglsk
 181 ggfvpwf
```

His29Ala/Ser50Cys

```
   1 mdhyldirlr pdpefppaql msvlfgkl**a**q alvaqggdri gvsfpdlde**c** rsrlgerlri
  61 hasaddlral larpwleglr dhlqfgepav vphptpyrqv srvqvksnpe rlrrrlmrrh
 121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfrlfir hgplqataee ggftcyglsk
 181 ggfvpwf
```

FIG. 39

Secuencias de Csy4

<213> Pseudomonas aeruginosa

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala Leu
Val Ala Gln Gly Gly Asp Arg Ile Gly val Ser Phe Pro Asp Leu Asp
Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg
Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr Pro
Tyr Arg Gln Val Ser Arg Val Gln Val Lys Ser Asn Pro Glu Arg Leu
Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala Arg
Lys Arg Ile Pro Asp Thr val Ala Arg Ala Leu Asp Leu Pro Phe Val
Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg
His Gly Pro Leu Gln val Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe

<213> Pseudomonas aeruginosa

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala Leu
Val Ala Gln Gly Gly Asp Arg Ile Gly val Ser Phe Pro Asp Leu Asp
Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg
Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr Pro
Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg Leu
Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala Arg
Lys Arg Ile Pro Asp Thr Val Ala Arg Thr Leu Asp Leu Pro Phe Val
Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg
His Gly Pro Leu Gln Ala Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe

<213> Pseudomonas aeruginosa

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala Leu
Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu Asp
Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
Asp Asp Leu His Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg
Asp His Leu Gln Phe Gly Glu Ala Ala Val Val Pro His Pro Thr Pro
Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg Leu
Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala Arg
Lys Arg Ile Pro Asp Thr Val Ala Arg Thr Leu Asp Leu Pro Phe Val
Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg
His Gly Pro Leu Gln Ala Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
Gly Leu Ser Lys Gly Gly Phe val Pro Trp Phe

FIG. 40

Secuencias de Cas6

Thermus thermophilus HB8

Met Val Leu Ala Ala Leu Val Leu Val Leu Glu Gly Glu Gly Leu Pro
Glu Pro Leu Gly Leu Arg Gly Phe Phe Tyr Gly Leu Leu Arg Glu Val
Ala Pro Glu Val His Asp Gln Gly Glu Asn Pro Phe Ala Leu Gly Phe
Gly Gly Arg Glu Gly Ala Ala Trp Ala Arg Val Ser Leu Leu Val Glu
Gly Leu Tyr Ala Arg Leu Ala Pro Arg Leu Tyr Ala Leu Glu Gly Glu
Glu Val Arg Leu Gly Pro Pro Phe Arg Val Arg Ala Val Leu Gln Glu
Gly His Pro Trp Ala Gly Val Ser Thr Tyr Pro Arg Leu Phe Gln Gly
Pro Pro Ser Arg Asp Leu Ala Leu Arg Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly Val His Tyr Pro Val Pro Glu Pro Arg Leu Val Leu
Glu Ser Leu Leu Arg Arg Leu Glu Ala Phe Gly Pro Leu Lys Ala Pro
Glu Gly Val Arg Glu Ala Leu Leu Glu Arg Thr Thr Val Arg Ser Leu
Glu Gly Arg Thr Leu Pro Ala Arg Thr Glu Val Asp Thr Ala Gly Phe
Val Gly Arg Val Val Tyr His Leu Pro Arg Ala Thr Glu Glu Glu Ala
Leu Trp Leu Ser Ala Leu Gly Arg Phe Ala Phe Tyr Ser Gly Val Gly
Ala Lys Thr Ser Leu Gly Tyr Gly Arg Ala Arg Ala Glu Ser Ala

Thermus thermophilus HB8

Met Pro Gln Ala Val Val Leu Glu Leu Val Gly Glu Lys Pro Pro Leu
Tyr Pro Ala Arg Tyr Ala His Gly Leu Phe Phe Ala Leu Leu Ser Arg
Val Ser Pro Glu Leu Ala Gln Lys Leu His Glu Ala Pro Arg Lys Pro
Phe Thr Leu Ala Pro Leu Pro Arg Ala Gly Pro Glu Gly Ala Thr Leu
Lys Gly Thr Leu Arg Leu Arg Leu Thr Thr Leu Asp Asp Gly Leu Phe
Ala Pro Phe Leu Arg Ala Leu Leu Glu Ala Ala Pro Asp Gly Leu Pro
Leu Gly Asp Ser Ser Tyr Arg Leu Ala Arg Val Leu Ala Thr Arg Glu
Gly His Pro Leu Ala Gly Ala Thr Ser Trp Glu Glu Leu Lys Glu Ala
Pro Lys Arg Glu Lys Ala Thr Phe Arg Phe Leu Thr Pro Thr Val Phe
Ala Thr Ser Lys Pro Gly Gly Arg Thr Arg Tyr Thr Pro Leu Pro Asp
Pro Arg Leu Ile Ala Gly Ser Leu Leu Asp Lys Trp Gln Ala His Ser
Pro Phe Pro Tyr Asn Pro Lys Glu Glu Ala Ala Leu Arg Glu Leu Phe
Glu Leu Asp Leu Glu Val Ala Gly Phe Arg Asn Leu Arg Phe His Arg
Val Gln Ala Gly Lys Gly Phe Phe Pro Gly Phe Thr Gly Glu Ala Thr
Leu Arg Leu Trp Ser Gln Ser Leu Glu Ala Gln Glu Ala Leu Gly Arg
Leu His Ala Leu Ala Phe Phe Ser Gly Val Gly Ala Lys Thr Pro Tyr
Gly Met Gly Leu Ala Val Pro Leu

FIG. 41A

Secuencias de Cas6

Staphylococcus epidermidis RP62A

Met Ile Asn Lys Ile Thr Val Glu Leu Asp Leu Pro Glu Ser Ile Arg
Phe Gln Tyr Leu Gly Ser Val Leu His Gly Val Leu Met Asp Tyr Leu
Ser Asp Asp Ile Ala Asp Gln Leu His His Glu Phe Ala Tyr Ser Pro
Leu Lys Gln Arg Ile Tyr His Lys Asn Lys Lys Ile Ile Trp Glu Ile
Val Cys Met Ser Asp Asn Leu Phe Lys Glu Val Val Lys Leu Phe Ser
Ser Lys Asn Ser Leu Leu Leu Lys Tyr Tyr Gln Thr Asn Ile Asp Ile
Gln Ser Phe Gln Ile Glu Lys Ile Asn Val Gln Asn Met Met Asn Gln
Leu Leu Gln Val Glu Asp Leu Ser Arg Tyr Val Arg Leu Asn Ile Gln
Thr Pro Met Ser Phe Lys Tyr Gln Asn Ser Tyr Met Ile Phe Pro Asp
Val Lys Arg Phe Phe Arg Ser Ile Met Ile Gln Phe Asp Ala Phe Phe
Glu Glu Tyr Arg Met Tyr Asp Lys Glu Thr Leu Asn Phe Leu Glu Lys
Asn Val Asn Ile Val Asp Tyr Lys Leu Lys Ser Thr Arg Phe Asn Leu
Glu Lys Val Lys Ile Pro Ser Phe Thr Gly Glu Ile Val Phe Lys Ile
Lys Gly Pro Leu Pro Phe Leu Gln Leu Thr His Phe Leu Leu Lys Phe
Gly Glu Phe Ser Gly Ser Gly Ile Lys Thr Ser Leu Gly Met Gly Lys
Tyr Ser Ile Ile

Mycobacterium tuberculosis H37Rv

Met Ala Ala Arg Arg Gly Gly Ile Arg Arg Thr Asp Leu Leu Arg Arg
Ser Gly Gln Pro Arg Gly Arg His Arg Ala Ser Ala Ala Glu Ser Gly
Leu Thr Trp Ile Ser Pro Thr Leu Ile Leu Val Gly Phe Ser His Arg
Gly Asp Arg Arg Met Thr Glu His Leu Ser Arg Leu Thr Leu Thr Leu
Glu Val Asp Ala Pro Leu Glu Arg Ala Arg Val Ala Thr Leu Gly Pro
His Leu His Gly Val Leu Met Glu Ser Ile Pro Ala Asp Tyr Val Gln
Thr Leu His Thr Val Pro Val Asn Pro Tyr Ser Gln Tyr Ala Leu Ala
Arg Ser Thr Thr Ser Leu Glu Trp Lys Ile Ser Thr Leu Thr Asn Glu
Ala Arg Gln Gln Ile Val Gly Pro Ile Asn Asp Ala Ala Phe Ala Gly
Phe Arg Leu Arg Ala Ser Gly Ile Ala Thr Gln Val Thr Ser Arg Ser
Leu Glu Gln Asn Pro Leu Ser Gln Phe Ala Arg Ile Phe Tyr Ala Arg
Pro Glu Thr Arg Lys Phe Arg Val Glu Phe Leu Thr Pro Thr Ala Phe
Lys Gln Ser Gly Glu Tyr Val Phe Trp Pro Asp Pro Arg Leu Val Phe
Gln Ser Leu Ala Gln Lys Tyr Gly Ala Ile Val Asp Gly Glu Glu Pro
Asp Pro Gly Leu Ile Ala Glu Phe Gly Gln Ser Val Arg Leu Ser Ala
Phe Arg Val Ala Ser Ala Pro Phe Ala Val Gly Ala Ala Arg Val Pro
Gly Phe Thr Gly Ser Ala Thr Phe Thr Val Arg Gly Val Asp Thr Phe
Ala Ser Tyr Ile Ala Ala Leu Leu Trp Phe Gly Glu Phe Ser Gly Cys
Gly Ile Lys Ala Ser Met Gly Met Gly Ala Ile Arg Val Gln Pro Leu
Ala Pro Arg Glu Lys Cys Val Pro Lys Pro

FIG. 41B

Secuencias de Cas6

Streptococcus thermophilus LMG 18311

Met Lys Lys Leu Val Phe Thr Phe Lys Arg Ile Asp His Pro Ala Gln
Asp Leu Ala Val Lys Phe His Gly Phe Leu Met Glu Gln Leu Asp Ser
Asp Tyr Val Asp Tyr Leu His Gln Gln Gln Thr Asn Pro Tyr Ala Thr
Lys Val Ile Gln Gly Lys Glu Asn Thr Gln Trp Val Val His Leu Leu
Thr Asp Asp Ile Glu Asp Lys Val Phe Met Thr Leu Leu Gln Ile Lys
Glu Val Ser Leu Asn Asp Leu Pro Lys Leu Ser Val Glu Lys Val Glu
Ile Gln Glu Leu Gly Ala Asp Lys Leu Leu Glu Ile Phe Asn Ser Glu
Glu Asn Gln Thr Tyr Phe Ser Ile Ile Phe Glu Thr Pro Thr Gly Phe
Lys Ser Gln Gly Ser Tyr Val Ile Phe Pro Ser Met Arg Leu Ile Phe
Gln Ser Leu Met Gln Lys Tyr Gly Arg Leu Val Glu Asn Gln Pro Glu
Ile Glu Glu Asp Thr Leu Asp Tyr Leu Ser Glu His Ser Thr Ile Thr
Asn Tyr Arg Leu Glu Thr Ser Tyr Phe Arg Val His Arg Gln Arg Ile
Pro Ala Phe Arg Gly Lys Leu Thr Phe Lys Val Gln Gly Ala Gln Thr
Leu Lys Ala Tyr Val Lys Met Leu Leu Thr Phe Gly Glu Tyr Ser Gly
Leu Gly Met Lys Thr Ser Leu Gly Met Gly Gly Ile Lys Leu Glu Glu
Arg Lys Asp

Streptococcus sanguinis SK36

Met Lys Lys Ile Arg Leu His Leu Ser Lys Val Ser Leu Lys Asp Asp
Asp Leu Val Cys Lys Leu Gln Gly Phe Leu Met Glu Lys Leu Ser Asp
Asp Phe Ala Ser Phe Leu His Gln Gln Glu Thr Asn Pro Tyr Ser Met
Asn Leu Arg Ser Glu Arg Glu Glu Ser Ile Trp Thr Val Asn Leu Leu
Ser Glu Glu Ala Glu Gln Gln Ile Leu Pro Gln Leu Leu Ser Leu Glu
Met Ile Lys Leu Glu Thr Tyr Ser Glu Glu Ile Leu Val Lys Asn Ile
Glu Ile Gln Ser Leu Ser Ser Gln Ser Leu Leu Glu Val Phe Gln Gly
Asp Glu Ala Ser His Leu Ile Ser Leu Asn Phe Tyr Thr Pro Thr Thr
Phe Lys Arg Gln Gly Gln Phe Val Leu Phe Pro Asp Thr Arg Leu Ile
Phe Gln Ser Leu Met Gln Lys Tyr Ser Arg Leu Val Glu Gly Lys Ala
Glu Ile Glu Glu Glu Thr Leu Glu Phe Leu Ala Glu His Ser Gln Ile
Ser Ser Tyr Arg Leu Lys Ser His Tyr Phe Pro Ile His Gly Arg Lys
Tyr Pro Ala Phe Glu Gly Arg Val Thr Ile Arg Ile Gln Gly Ala Ser
Thr Leu Lys Ala Tyr Ala Gln Met Leu Leu Arg Phe Gly Glu Tyr Ser
Gly Val Gly Ala Lys Cys Ser Leu Gly Met Gly Gly Met Arg Ile Glu
Glu Arg Lys Thr

FIG. 41C

Secuencias de Cas6

Microcystis aeruginosa NIES-843

Met Pro Tyr Ser Leu Val Leu Asn Leu Thr Pro Arg Ser Pro Ile Tyr
Pro Asn Phe Leu Thr Gly Arg His Leu His Ala Leu Phe Leu Thr Leu
Val Ser Ser Val Asp Gln Glu Leu Gly Lys Ile Leu His Thr Ala Glu
Ala Asp Lys Ala Phe Thr Leu Ser Pro Leu Gln Met Gln Ser Gly Gly
Lys Thr Ile Asn Ser Pro Gln Trp Arg Tyr Glu Arg Pro Ile Ala Pro
Glu Thr Pro Cys Trp Trp Arg Ile Ser Leu Leu Asp Asp Arg Leu Phe
Gly Lys Leu Thr Pro Leu Trp Leu Asn Leu Asn Pro Lys His Pro Trp
His Leu Gly Ser Ala Asp Leu Val Ile Thr Ser Val Leu Ala Thr Pro
Gln Ser Val Gln Pro Trp Ala Asn Ser Cys Thr Tyr Gln Tyr Leu Tyr
Glu Asn Ala Ser Glu Thr Asn Arg Glu Phe Asp Phe Leu Phe Ala Thr
Pro Val Thr Phe Arg Gln Gly Lys Phe Asp Ser Ala Leu Pro Thr Arg
Glu Leu Val Phe Asn Ser Leu Leu Asn Arg Trp Asn Arg Tyr Ser Ala
Ile Pro Phe Asp Ser Ile Val Leu Glu Ser Ile Phe Pro Ser Phe Phe
Asp Ile Gln Thr Lys Leu Ala Asp Glu Ala Tyr Lys Asn Gln Ser Phe
Gly Cys Val Gly Glu Ile His Tyr Arg Leu Leu Gly Glu Val Glu Pro
Ala Lys Ile Lys Ala Ile Asn Val Leu Ala Asp Phe Ala Leu Tyr Ala
Gly Val Gly Arg Lys Thr Thr Met Gly Met Gly Met Thr Arg Arg Ile
Ala Lys Glu Lys Arg

Meiothermus silvanus DSM 9946

Met Met Leu Ala Ala Leu Val Leu Pro Leu Glu Gly Gln Ala Arg Pro
Asp Pro Asp Gly Trp Arg Gly Leu Val Tyr Gly Leu Leu Lys Glu Ile
Asp Pro Glu Leu His Thr Ala Gln His Asn Pro Phe Ser Leu Gly Leu
Gly Gly Ala Glu Gly Gln Trp Trp Val Arg Ile Ala Leu Leu Glu Glu
Gly Leu Tyr Ala Arg Leu Ser Pro His Leu Phe Gly Leu Val Gly Gln
Ser Val Lys Leu Lys Glu Pro Phe Arg Val Arg Ala Val Leu Gln Glu
Glu His Pro Trp Ala Ser Leu Ser Thr Tyr Pro Arg Leu Phe Gln Gly
Gln Ala Ser Pro Ser Leu Gly Leu Gln Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly Asn Ser Tyr Pro Leu Pro Glu Pro Lys Leu Val Phe
Asp Ser Leu Thr Gln Arg Trp Asn Ala Phe Ala Pro Val Lys Val Pro
Pro Glu Met Ala Glu Thr Trp Glu Arg Val Thr Ile Thr Arg Leu Gln
Gly His Thr Gln Ala Ile Arg Pro Asn Pro Asp Glu Arg Gly Val Gly
Phe Val Gly Arg Val Val Tyr His Leu Pro Ala Ala Lys Pro Thr Glu
Ala Gln Trp Met Gln Ala Leu Gly Arg Phe Ala Phe Tyr Ala Gly Val
Gly Ala Lys Thr Ser Leu Gly Phe Gly Arg Val Arg Gly Phe Asp Pro
Ile Leu Lys Glu Glu Ser Ala Asn Gly Arg Leu Asp Ala Glu Asp Ser
Ser Ser Leu Ala Thr Pro Gln Asp Pro Gly Ala

FIG. 41D

Secuencias de Cas6

Meiothermus silvanus DSM 9946

Met Met Leu Ala Ala Leu Val Leu Pro Leu Glu Gly Gln Ala Arg Pro
Asp Pro Asp Gly Trp Arg Gly Leu Val Tyr Gly Leu Leu Lys Glu Ile
Asp Pro Glu Leu His Thr Ala Gln His Asn Pro Phe Ser Leu Gly Leu
Gly Gly Ala Glu Gly Gln Trp Trp Val Arg Ile Ala Leu Leu Glu Glu
Gly Leu Tyr Ala Arg Leu Ser Pro His Leu Phe Gly Leu Val Gly Gln
Ser Val Lys Leu Lys Glu Pro Phe Arg Val Arg Ala Val Leu Gln Glu
Glu His Pro Trp Ala Ser Leu Ser Thr Tyr Pro Arg Leu Phe Gln Gly
Gln Ala Ser Pro Ser Leu Gly Leu Gln Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly Asn Ser Tyr Pro Leu Pro Glu Pro Lys Leu Val Phe
Asp Ser Leu Thr Gln Arg Trp Asn Ala Phe Ala Pro Val Lys Val Pro
Pro Glu Met Ala Glu Thr Trp Glu Arg Val Thr Ile Thr Arg Leu Gln
Gly His Thr Gln Ala Ile Arg Pro Asn Pro Asp Glu Arg Gly Val Gly
Phe Val Gly Arg Val Val Tyr His Leu Pro Ala Ala Lys Pro Thr Glu
Ala Gln Trp Met Gln Ala Leu Gly Arg Phe Ala Phe Tyr Ala Gly Val
Gly Ala Lys Thr Ser Leu Gly Phe Gly Arg Val Arg Gly Phe Asp Pro
Ile Leu Lys Glu Glu Ser Ala Asn Gly Arg Leu Asp Ala Glu Asp Ser
Ser Ser Leu Ala Thr Pro Gln Asp Pro Gly Ala

Meiothermus ruber DSM 1279

Met Ile Leu Ala Ala Leu Ile Leu Pro Leu Glu Gly Pro Thr Arg Pro
Asp Pro Asp Gly Trp Arg Gly Leu Val Tyr Gly Leu Leu Lys Glu Ile
Asp Pro Glu Leu His Ala Ala Gln His Asn Pro Phe Ser Leu Gly Leu
Gly Gly Ala Leu Gly Gln Trp Trp Val Arg Ile Ala Phe Leu Glu Glu
Gly Leu Tyr Ala Arg Leu Ser Pro His Leu Phe Gly Leu Ala Gly Gln
Thr Val Arg Leu Lys Glu Ala Phe Gln Val Arg Ala Val Leu Gln Glu
Ala His Pro Trp Ala Gly Val Ser Thr Tyr Pro Lys Leu Phe Gln Gly
Gln Ala Thr Ala Ser Leu Gly Leu Gln Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly His Ser Tyr Pro Leu Pro Glu Pro Arg Leu Val Phe
Glu Ser Leu Thr Gln Arg Trp Asn Ala Phe Ala Pro Val Lys Val Pro
Gln Glu Val Gln Glu Ala Trp Glu Arg Leu Leu Val Gly Gln Phe Gln
Gly Arg Thr His His Ile Ala Pro Asn Gln Asp Glu Arg Gly Val Gly
Phe Val Gly Arg Val Val Tyr Tyr Leu Pro Lys Ala Ser Pro Thr Glu
Ala Gln Trp Leu Gln Ala Leu Gly Arg Phe Ala Phe Tyr Ala Gly Val
Gly Ala Lys Thr Ser Leu Gly Phe Gly Arg Val Arg Met Phe Asp Pro
Leu Gln Gln Glu Arg Arg Pro Asp Glu Ser Glu Gln Gly Ala Leu Thr
Gly Thr Val Gly Gly Val

FIG. 41E

Secuencias de Cas6

Thermus thermophilus HB8

Met Val Leu Ala Ala Leu Val Leu Val Leu Glu Gly Glu Gly Leu Pro
Glu Pro Leu Gly Leu Arg Gly Phe Phe Tyr Gly Leu Leu Arg Glu Val
Ala Pro Glu Val His Asp Gln Gly Glu Asn Pro Phe Ala Leu Gly Phe
Gly Gly Arg Glu Gly Ala Ala Trp Ala Arg Val Ser Leu Leu Val Glu
Gly Leu Tyr Ala Arg Leu Ala Pro Arg Leu Tyr Ala Leu Glu Gly Glu
Glu Val Arg Leu Gly Pro Pro Phe Arg Val Arg Ala Val Leu Gln Glu
Gly His Pro Trp Ala Gly Val Ser Thr Tyr Pro Arg Leu Phe Gln Gly
Pro Pro Ser Arg Asp Leu Ala Leu Arg Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly Val His Tyr Pro Val Pro Glu Pro Arg Leu Val Leu
Glu Ser Leu Leu Arg Arg Leu Glu Ala Phe Gly Pro Leu Lys Ala Pro
Glu Gly Val Arg Glu Ala Leu Leu Glu Arg Thr Thr Val Arg Ser Leu
Glu Gly Arg Thr Leu Pro Ala Arg Thr Glu Val Asp Thr Ala Gly Phe
Val Gly Arg Val Val Tyr His Leu Pro Arg Ala Thr Glu Glu Glu Ala
Leu Trp Leu Ser Ala Leu Gly Arg Phe Ala Phe Tyr Ser Gly Val Gly
Ala Lys Thr Ser Leu Gly Tyr Gly Arg Ala Arg Ala Glu Ser Ala

Thermus thermophilus JL-18

Met Val Leu Ala Ala Leu Val Leu Val Leu Glu Gly Glu Gly Leu Pro
Glu Pro Leu Gly Leu Arg Gly Phe Phe Tyr Gly Leu Leu Arg Glu Val
Ala Pro Glu Val His Asp Gln Gly Glu Asn Pro Phe Ala Leu Gly Phe
Gly Gly Arg Glu Gly Ala Ser Trp Ala Arg Val Ser Leu Leu Arg Glu
Glu Leu Tyr Ala Arg Leu Ala Pro Arg Leu Tyr Ala Leu Glu Gly Glu
Glu Val Arg Leu Gly Pro Pro Phe Arg Val Arg Ala Val Leu Gln Glu
Gly His Pro Trp Ala Gly Val Ser Thr Tyr Pro Arg Leu Phe Gln Gly
Pro Pro Ser Arg Asp Leu Ala Leu Arg Phe Ala Ser Pro Thr Phe Phe
Arg Arg Lys Gly Val His Tyr Pro Val Pro Glu Pro Arg Leu Val Leu
Glu Ser Leu Leu Arg Arg Leu Glu Ala Phe Gly Pro Leu Lys Ala Pro
Glu Gly Val Arg Glu Ala Leu Leu Glu Arg Thr Thr Val Arg Ser Leu
Glu Gly Arg Thr Leu Pro Ala Arg Thr Glu Val Asp Thr Ala Gly Phe
Val Gly Arg Val Val Tyr His Leu Pro Arg Ala Thr Glu Glu Glu Ala
Leu Trp Leu Ser Ala Leu Gly Arg Phe Ala Phe Tyr Ser Gly Val Gly
Ala Lys Thr Ser Leu Gly Tyr Gly Arg Ala Arg Ala Glu Ser Pro

FIG. 41F

Secuencias de Cas6

Thermus sp. CCB US3 UF1

Met Leu Ala Ala Leu Val Leu Thr Leu Glu Gly Glu Ala Pro Pro Glu
Pro Arg Gly Leu Arg Gly Phe Phe Tyr Gly Leu Leu Gln Glu Val Ala
Pro Glu Val His Asp Gln Gly Glu Asn Pro Phe Ala Leu Gly Phe Gly
Gly Lys Glu Gly Ala Tyr Trp Ala Arg Phe Ser Leu Leu Gln Glu Gly
Leu Tyr Ala Arg Leu Ala Pro Arg Leu Phe Ala Leu Glu Gly Lys Glu
Val Arg Leu Gly Lys Pro Phe Arg Val Arg Gly Val Leu Gln Glu Gly
His Pro Trp Ala Gly Val Ser Thr Tyr Ala Arg Leu Phe Gln Gly Glu
Ala Leu Pro Asp Leu Pro Leu Arg Phe Ala Ser Pro Thr Phe Phe Arg
Arg Lys Gly Val His Tyr Pro Leu Pro Glu Pro Arg Leu Val Val Glu
Ser Leu Leu Arg Arg Leu Glu Ala Phe Gly Pro Leu Lys Ala Pro Glu
Gly Val Arg Glu Ala Leu Leu Glu Arg Thr Thr Val Arg Trp Phe Glu
Gly Lys Thr Leu Lys Ala Glu Thr Glu Val Glu Ala Val Gly Phe Val
Gly Lys Val Val Tyr His Leu Pro Arg Ala Thr Glu Glu Glu Ala Arg
Trp Leu Gln Ala Leu Gly Arg Phe Ala Phe Tyr Ser Gly Val Gly Ala
Lys Thr Gly Leu Gly Tyr Gly Arg Ala Arg Val Gly

Thermus aquaticus Y51MC23

Met Val Leu Val Ala Leu Val Leu Val Leu Glu Gly Glu Gly Pro Pro
Glu Pro Leu Gly Leu Arg Gly Phe Phe Tyr Thr Leu Leu Lys Glu Ala
Phe Pro Glu Leu His Asp Gln Gly Glu Asn Pro Phe Ala Leu Gly Phe
Gly Leu Arg Gly Gly Glu Pro Trp Ala Arg Val Ser Leu Leu Arg Glu
Asp Leu Tyr Gly Arg Leu Ser Pro Ala Leu Phe Gly Leu Glu Gly Arg
Glu Val Arg Leu Gly Arg Leu Phe Arg Val Arg Ala Val Leu Gln Glu
Gly His Pro Trp Ala Gly Leu Thr Thr Tyr Ala Arg Leu Phe Gln Gly
Pro His Ser Pro Asn Leu Pro Leu Arg Phe Tyr Ser Pro Thr Phe Phe
Arg Arg Lys Gly Val Gln Tyr Pro Leu Pro Glu Pro Arg Leu Val Leu
Glu Ser Leu Leu Arg Arg Leu Glu Ala Phe Gly Pro Leu Lys Ala Pro
Gln Glu Val Arg Glu Ala Leu Leu Glu Arg Thr Thr Val Arg Phe Leu
Glu Gly Arg Thr Gln Met Ala Arg Thr Glu Val Asp Thr Val Gly Phe
Val Gly Lys Val Val Tyr His Leu Pro Lys Ala Thr Glu Glu Glu Ala
Leu Trp Leu Ser Ala Leu Gly Arg Tyr Ala Phe Phe Ser Gly Val Gly
Ala Lys Thr Ser Leu Gly Tyr Gly Leu Ala Arg Ala Phe Thr Gln Val
Gly Pro Gln Asp Ala Glu Thr

FIG. 41G

Secuencias de Cas6

Marinithermus hydrothermalis DSM 14884

Met Leu Leu Ala Ala Leu Val Leu Pro Leu Glu Gly Pro Asp Arg Pro
Gln Pro Leu His Ala Arg Gly Trp Val Tyr Arg Leu Leu Arg Glu Ala
Ala Pro Glu Ile His Asp Ala Glu Gly Pro Lys Pro Phe Thr Val Gly
Val Gly Gly Arg Pro Asn Ala Val Trp Val Arg Leu Thr Cys Leu Ala
Glu Glu Val Tyr Ala Ala Leu Ser Pro Arg Leu Trp Ser Gln Val Gly
Leu Glu Val Arg Leu Gly Glu Asp Thr Tyr Arg Ile Lys Ala Val Leu
Glu Ala Glu His Pro Trp Ala Gly Leu Ala Thr Trp Pro Arg Leu Phe
Gln Gly Glu Ala Gly Pro Asp Leu Gly Leu Glu Phe Ala Ser Pro Thr
Phe Phe Arg Arg Gln Gly Ala Asn Tyr Pro Leu Pro Glu Pro Arg Leu
Val Leu Gly Ser Leu Ile Glu Arg Trp Asn Ala His Ala Pro Thr Pro
Val Pro Pro Glu Val Ala Glu Arg Leu Val Glu Ala Thr Thr Leu Arg
Tyr Leu Lys Gly His Thr Val Ser Ala Val Gly His Asp Arg Thr Val
Gly Phe Arg Gly Arg Val Thr Tyr His Leu Pro Arg Ala Ser Thr Glu
Glu Ala Arg Trp Leu Ala Ala Leu Gly Arg Phe Ala Phe Phe Ser Gly
Val Gly Ala Lys Thr Thr Leu Gly Phe Gly Gln Val Arg Pro Tyr Pro
Leu Leu Ala Pro Ser Ala Ala Pro Pro Gly Pro

Thermus thermophilus HB8

Met Pro Gln Ala Val Val Leu Glu Leu Val Gly Glu Lys Pro Pro Leu
Tyr Pro Ala Arg Tyr Ala His Gly Leu Phe Phe Ala Leu Leu Ser Arg
Val Ser Pro Glu Leu Ala Gln Lys Leu His Glu Ala Pro Arg Lys Pro
Phe Thr Leu Ala Pro Leu Pro Arg Ala Gly Pro Glu Gly Ala Thr Leu
Lys Gly Thr Leu Arg Leu Arg Leu Thr Thr Leu Asp Asp Gly Leu Phe
Ala Pro Phe Leu Arg Ala Leu Leu Glu Ala Ala Pro Asp Gly Leu Pro
Leu Gly Asp Ser Ser Tyr Arg Leu Ala Arg Val Leu Ala Thr Arg Glu
Gly His Pro Leu Ala Gly Ala Thr Ser Trp Glu Glu Leu Lys Glu Ala
Pro Lys Arg Glu Lys Ala Thr Phe Arg Phe Leu Thr Pro Thr Val Phe
Ala Thr Ser Lys Pro Gly Gly Arg Thr Arg Tyr Thr Pro Leu Pro Asp
Pro Arg Leu Ile Ala Gly Ser Leu Leu Asp Lys Trp Gln Ala His Ser
Pro Phe Pro Tyr Asn Pro Lys Glu Glu Ala Ala Leu Arg Glu Leu Phe
Glu Leu Asp Leu Glu Val Ala Gly Phe Arg Asn Leu Arg Phe His Arg
Val Gln Ala Gly Lys Gly Phe Phe Pro Gly Phe Thr Gly Glu Ala Thr
Leu Arg Leu Trp Ser Gln Ser Leu Glu Ala Gln Glu Ala Leu Gly Arg
Leu His Ala Leu Ala Phe Phe Ser Gly Val Gly Ala Lys Thr Pro Tyr
Gly Met Gly Leu Ala Val Pro Leu

FIG. 41H

Secuencias de Cas6

Thermus thermophilus HB27

Met Pro Gln Ala Val Val Leu Glu Leu Val Gly Glu Lys Pro Pro Leu
Tyr Pro Ala Arg Tyr Ala His Gly Leu Phe Phe Ala Leu Leu Ser Arg
Val Ser Pro Glu Leu Ala Gln Lys Leu His Glu Ala Pro Arg Lys Pro
Phe Thr Leu Ala Pro Leu Pro Arg Ala Gly Pro Glu Gly Ala Thr Leu
Lys Gly Thr Leu Arg Leu Arg Leu Thr Thr Leu Asp Asp Gly Leu Phe
Ala Pro Phe Leu Arg Ala Leu Leu Glu Ala Ala Pro Asp Gly Leu Pro
Leu Gly Asp Ser Ser Tyr Arg Leu Ala Arg Val Leu Ala Thr Arg Glu
Gly His Pro Leu Ala Gly Ala Thr Ser Trp Glu Glu Leu Lys Glu Ala
Pro Lys Arg Glu Lys Val Thr Phe Arg Phe Leu Thr Pro Thr Val Phe
Ala Thr Ser Lys Pro Gly Gly Arg Thr Arg Tyr Thr Pro Leu Pro Asp
Pro Arg Leu Ile Ala Gly Ser Leu Leu Asp Lys Trp Gln Ala His Ser
Pro Phe Pro Tyr Asn Pro Lys Glu Glu Ala Ala Leu Arg Gly Leu Phe
Glu Leu Asp Leu Glu Val Ala Gly Phe Arg Asn Leu Arg Phe His Arg
Val Gln Ala Gly Lys Gly Phe Phe Pro Gly Phe Thr Gly Glu Met Thr
Leu Arg Leu Trp Ser Gln Ser Leu Glu Ala Arg Glu Ala Leu Gly Arg
Leu His Ala Leu Ala Phe Phe Ser Gly Val Gly Ala Lys Thr Pro Tyr
Gly Met Gly Leu Ala Val Pro Leu

Thermus thermophilus SG0.5JP17-16

Met Pro Gln Ala Val Val Leu Glu Leu Val Gly Glu Lys Pro Pro Leu
Tyr Pro Gly Arg Tyr Ala His Gly Leu Phe Phe Ala Leu Leu Ser Arg
Val Ser Pro Glu Leu Ala Gln Lys Leu His Glu Ala Pro Arg Lys Pro
Phe Thr Leu Ala Pro Leu Pro Arg Val Gly Pro Glu Gly Ala Thr Leu
Lys Gly Ile Leu Arg Leu Arg Leu Thr Ala Leu Asp Asp Gly Leu Phe
Ala Pro Phe Leu Arg Ala Leu Leu Glu Ala Ala Pro Asp Gly Leu Pro
Leu Gly Asp Ser Ser Tyr Arg Leu Ala Arg Val Leu Ala Thr Arg Glu
Gly His Pro Leu Ala Gly Ala Thr Ser Trp Glu Glu Leu Lys Glu Ala
Pro Lys Arg Glu Lys Ala Thr Phe Arg Phe Leu Thr Pro Thr Val Phe
Ala Thr Ser Lys Pro Gly Gly Arg Thr Arg Tyr Thr Pro Leu Pro Asp
Pro Arg Leu Ile Ala Gly Ser Leu Leu Asp Lys Trp Gln Ala His Ser
Pro Phe Pro Tyr Asn Pro Lys Glu Glu Ala Ala Leu Arg Glu Leu Phe
Glu Leu Asp Leu Glu Val Ala Gly Phe Arg Asn Leu Arg Phe His Arg
Val Gln Ala Gly Lys Gly Phe Phe Pro Gly Phe Thr Gly Glu Ala Thr
Leu Arg Leu Trp Ser Gln Ser Leu Glu Ala Gln Glu Ala Leu Gly Arg
Leu His Ala Leu Ala Phe Phe Ser Gly Val Gly Ala Lys Thr Pro Tyr
Gly Met Gly Leu Ala Val Pro Leu

FIG. 41I

Secuencias de Cas6

Thermus thermophilus JL-18

Met Pro Gln Ala Val Val Leu Glu Leu Val Gly Glu Glu Ser Pro Leu
Tyr Pro Ala Arg Tyr Ala His Gly Leu Phe Phe Ala Leu Leu Ser Arg
Val Ser Pro Glu Leu Ala Gln Lys Leu His Glu Ala Pro Arg Lys Pro
Phe Thr Leu Ala Pro Leu Pro Arg Val Gly Ser Glu Gly Ala Thr Leu
Lys Gly Ile Leu Arg Leu Arg Leu Thr Ile Leu Asp Asp Gly Leu Phe
Ala Pro Phe Leu Arg Ala Leu Leu Glu Ala Ala Pro Asp Gly Leu Pro
Leu Gly Asp Ser Ser Tyr Arg Leu Ala Arg Val Leu Ala Thr Arg Glu
Gly His Pro Leu Ala Gly Ala Thr Ser Trp Glu Glu Leu Lys Glu Ala
Pro Lys Arg Glu Lys Ala Thr Phe Arg Phe Leu Thr Pro Thr Val Phe
Ala Thr Ser Lys Pro Gly Gly Arg Thr Arg Tyr Thr Pro Leu Pro Asp
Pro Arg Leu Ile Ala Gly Ser Leu Leu Asp Lys Trp Gln Ala His Ser
Pro Phe Pro Tyr Asn Pro Lys Glu Glu Ala Ala Leu Arg Glu Leu Phe
Glu Leu Asp Leu Glu Val Ala Gly Phe Arg Asn Leu Arg Phe His Arg
Val Gln Ala Gly Lys Ser Phe Phe Pro Gly Phe Thr Gly Glu Met Thr
Leu Arg Leu Trp Ser Gln Ser Leu Glu Ala Gln Gly Ala Leu Gly Arg
Leu His Ala Leu Ala Phe Phe Ser Gly Val Gly Ala Lys Thr Pro Tyr
Gly Met Gly Leu Ala Val Pro Leu

FIG. 41J

Secuencias de Cas5

Desulfovibrio vulgaris str. Hildenborough

Met Thr His Gly Ala Val Lys Thr Tyr Gly Ile Arg Leu Arg Val Trp
Gly Asp Tyr Ala Cys Phe Thr Arg Pro Glu Met Lys Val Glu Arg Val
Ser Tyr Asp Val Met Pro Pro Ser Ala Ala Arg Gly Ile Leu Glu Ala
Ile His Trp Lys Pro Ala Ile Arg Trp Ile Val Asp Arg Ile His Val
Leu Arg Pro Ile Val Phe Asp Asn Val Arg Arg Asn Glu Val Ser Ser
Lys Ile Pro Lys Pro Asn Pro Ala Thr Ala Met Arg Asp Arg Lys Pro
Leu Tyr Phe Leu Val Asp Asp Gly Ser Asn Arg Gln Gln Arg Ala Ala
Thr Leu Leu Arg Asn Val Asp Tyr Val Ile Glu Ala His Phe Glu Leu
Thr Asp Lys Ala Gly Ala Glu Asp Asn Ala Gly Lys His Leu Asp Ile
Phe Arg Arg Arg Ala Arg Ala Gly Gln Ser Phe Gln Gln Pro Cys Leu
Gly Cys Arg Glu Phe Pro Ala Ser Phe Glu Leu Leu Glu Gly Asp Val
Pro Leu Ser Cys Tyr Ala Gly Glu Lys Arg Asp Leu Gly Tyr Met Leu
Leu Asp Ile Asp Phe Glu Arg Asp Met Thr Pro Leu Phe Phe Lys Ala
Val Met Glu Asp Gly Val Ile Thr Pro Pro Ser Arg Thr Ser Pro Glu
Val Arg Ala

Neisseria mucosa ATCC 25996

Met Asn Gln Ile Arg Leu His Val Trp Gly Asp Tyr Ala Cys Phe Thr
Arg Pro Glu Met Lys Val Glu Arg Val Ser Tyr Asp Val Ile Thr Pro
Ser Ala Ala Arg Gly Ile Leu Ala Ala Val His Trp Lys Pro Ala Ile
Arg Trp Val Ile Asp Arg Ile Tyr Val Leu Lys Pro Ile Arg Phe Glu
Ser Val Arg Arg Asn Glu Leu Gly Gly Lys Ile Ser Ala Gly Lys Val
Ser Gly Ala Met Lys Arg Lys Ser Val Ala Asp Leu Tyr Thr Leu Ile
Glu Asp Asp Arg Gln Gln Arg Ala Ala Thr Val Leu Lys Asp Val Ala
Tyr Val Ile Glu Ala His Ala Val Leu Thr Ala Lys Ala Gly Ala Asp
Glu Thr Val Thr Lys His Ile Glu Met Phe Lys Arg Arg Ala Lys Lys
Gly Gln Cys Phe Gln Gln Pro Cys Leu Gly Val Arg Glu Phe Pro Ala
Asp Phe Ala Leu Ile Asp Glu Gly Glu Pro Leu Pro Pro Ser Ala Leu
Ser Glu Ser Glu Ala Asn Arg Asp Leu Gly Trp Met Leu His Asp Ile
Asp Phe Asp His Gly Asn Thr Pro His Phe Phe Arg Ala Gln Met Lys
Asp Gly Val Ile Asp Val Pro Pro Phe Tyr Ala Glu Glu Val Lys Ala

FIG. 42A

Secuencias de Cas5

Bacillus halodurans C-125

Met Arg Asn Glu Val Gln Phe Glu Leu Phe Gly Asp Tyr Ala Leu Phe
Thr Asp Pro Leu Thr Lys Ile Gly Gly Glu Lys Leu Ser Tyr Ser Val
Pro Thr Tyr Gln Ala Leu Lys Gly Ile Ala Glu Ser Ile Tyr Trp Lys
Pro Thr Ile Val Phe Val Ile Asp Glu Leu Arg Val Met Lys Pro Ile
Gln Met Glu Ser Lys Gly Val Arg Pro Ile Glu Tyr Gly Gly Gly Asn
Thr Leu Ala His Tyr Thr Tyr Leu Lys Asp Val His Tyr Gln Val Lys
Ala His Phe Glu Phe Asn Leu His Arg Pro Asp Leu Ala Phe Asp Arg
Asn Glu Gly Lys His Tyr Ser Ile Leu Gln Arg Ser Leu Lys Ala Gly
Gly Arg Arg Asp Ile Phe Leu Gly Ala Arg Glu Cys Gln Gly Tyr Val
Ala Pro Cys Glu Phe Gly Ser Gly Asp Gly Phe Tyr Asp Gly Gln Gly
Lys Tyr His Leu Gly Thr Met Val His Gly Phe Asn Tyr Pro Asp Glu
Thr Gly Gln His Gln Leu Asp Val Arg Leu Trp Ser Ala Val Met Glu
Asn Gly Tyr Ile Gln Phe Pro Arg Pro Glu Asp Cys Pro Ile Val Arg
Pro Val Lys Glu Met Glu Pro Lys Ile Phe Asn Pro Asp Asn Val Gln
Ser Ala Glu Gln Leu Leu His Asp Leu Gly Gly Glu

Bacillus cereus F65185

Met Lys Lys Glu Glu Glu Ser Leu Arg Asn Ser Ile Glu Phe Glu Val
Phe Gly Asp Tyr Ala Leu Phe Thr Asp Pro Leu Met Lys Met Gly Gly
Glu Lys Leu Thr Tyr Gln Val Pro Thr Tyr Gln Ala Ile Lys Gly Ile
Val Glu Ser Ile Tyr Trp Lys Pro Thr Leu Leu Met Ile Val Asp Lys
Ile Arg Ile Met Asn Ala Ile Lys Met Glu Ser Lys Gly Ile Arg Pro
Ile Glu Tyr Gly Gly Gly Asn Thr Leu Ala Asn Tyr Thr Tyr Leu Lys
Asn Val Arg Tyr Gln Val Gln Ala His Phe Ile Phe Asn Pro His Arg
Pro Asp Leu Ala Phe Asp Arg Asn Glu Tyr Lys His His Asn Ile Leu
Lys Arg Ser Leu Lys Val Gly Gly Arg Arg Asp Ile Phe Leu Gly Thr
Arg Glu Cys Gln Gly Tyr Val Glu Pro Cys Val Phe Gly Glu Gly Glu
Gly Phe Tyr Asp Asn Tyr Gly Gly Asp Ile His Leu Gly Thr Met Val
His Gly Leu Asn Tyr Pro Asp Glu Thr Gly Arg Asn Glu Leu Glu Val
Arg Leu Trp Asn Pro Val Met Arg Asp Gly Ile Ile Gln Phe Ile Arg
Pro Glu Glu Cys Thr Lys Ile Arg Lys Ile Ser Lys Met Glu Pro Lys
Ile Phe Asp Ser Ser Asn Val Glu Ser Val Asp Glu Leu Ile Lys Gln
Leu Glu Glu Gly Gly Glu

FIG. 42B

Secuencias de Cas5

Selenomonas noxia ATCC 43541

Met Arg Asn Ser Ile Glu Phe Gln Val Tyr Gly Arg Met Ala Leu Phe
Thr Asp Pro Ile Thr Lys Ile Gly Gly Glu Lys Ala Ser Tyr Ser Val
Pro Thr Tyr Gln Ala Leu Lys Gly Ile Thr Glu Ser Ile Tyr Trp Lys
Pro Thr Ile Ile Trp Phe Ile Asp Glu Val Arg Val Met Lys Arg Ile
Thr Thr Gln Val Arg Gly Val Lys Pro Leu Lys Tyr Gly Asp Ser Gly
Asn Asp Leu Ser Tyr Tyr Lys Tyr Leu Ser Asp Val Cys Tyr Gln Val
Arg Ala His Phe Glu Phe Asn Met His Arg Glu Glu Leu Lys Glu Asp
Arg Asp Glu His Lys His His Asn Ile Ala Lys Arg Met Val Glu Arg
Gly Gly Arg Arg Asp Ile Phe Leu Gly Thr Arg Glu Cys Gln Gly Tyr
Val Glu Pro Val Lys Tyr Gly Val Gly Lys Gly Tyr Tyr Asp Asn Val
Asp Glu Leu Pro Leu Gly Ile Met Leu His Gly Phe Asn Tyr Pro Asp
Glu Thr Gly Glu Asp Lys Leu Gln Val Arg Phe Trp Lys Pro Thr Met
Lys Lys Gly Ile Ile His Phe Arg Arg Pro Glu Lys Cys Glu Met Val
Arg Asp Ile Arg Glu Val Arg Thr Lys Gln Phe Asp Ala Asp Asn Val
Phe Phe Ala Glu Glu Glu Glu Lys Gln Leu Glu Gly Gly His Leu

FIG. 42C