

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 923 629**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 15/64** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.04.2020 PCT/US2020/026405**

87 Fecha y número de publicación internacional: **08.10.2020 WO20206134**

96 Fecha de presentación y número de la solicitud europea: **02.04.2020 E 20722744 (8)**

97 Fecha y número de publicación de la concesión europea: **15.06.2022 EP 3775201**

54 Título: **Métodos para la introducción sin cicatrices de modificaciones dirigidas en vectores de direccionamiento**

30 Prioridad:

**04.04.2019 US 201962829327 P**

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

**29.09.2022**

73 Titular/es:

**REGENERON PHARMACEUTICALS, INC.  
(100.0%)  
777 Old Saw Mill River Road  
Tarrytown, NY 10591, US**

72 Inventor/es:

**BRYDGES, SUSANNAH;  
ROJAS, JOSE F.;  
WARSHAW, GREGG S. y  
SIAO, CHIA-JEN**

74 Agente/Representante:

**SÁNCHEZ SILVA, Jesús Eladio**

ES 2 923 629 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Métodos para la introducción sin cicatrices de modificaciones dirigidas en vectores de direccionamiento

5 Referencia cruzada a las solicitudes relacionadas

Esta solicitud reivindica el beneficio de la solicitud de Estados Unidos núm. 62/829,327, presentada el 4 de abril de 2019.

10 Referencia a un listado de secuencias

Enviado como un archivo de texto a través de efs web

15 El Listado de Secuencias escrito en el archivo 544999SEQLIST.txt es de 20,7 kilobytes, fue creado el 21 de marzo de 2020.

Antecedentes

20 La construcción perfecta de ADN es de particular importancia cuando se crean líneas de animales transgénicos, ya que las cicatrices producidas por los sitios de restricción u otras manipulaciones pueden afectar negativamente la expresión génica si se producen en una región importante para la regulación. El direccionamiento al genoma de los mamíferos frecuentemente requiere la construcción de grandes vectores de direccionamiento con brazos de ADN largos para dirigir la recombinación homóloga, así como también casetes de resistencia a antibióticos para la selección de clones de células madre embrionarias. Los clones dirigidos correctamente contienen frecuentemente  
25 múltiples cicatrices necesarias para la construcción del vector, sin mencionar el casete de resistencia en sí. Para la ablación genética, estas lesiones pueden no ser importantes para el resultado final (un alelo nulo), pero siempre existe la posibilidad de que la expresión de los genes vecinos se vea afectada negativamente. Para modificaciones que no sean inactivación, tal como inserción, la expresión fiel del locus dirigido suele ser importante para los estudios en cuestión.

30 En particular, la humanización, el reemplazo directo de un gen de ratón con su homólogo humano, requiere uniones perfectas entre la secuencia de ratón y humana para que la maquinaria de transcripción del ratón replique fielmente la expresión del nuevo alelo. Se debe tener cuidado de ocultar las cicatrices de construcción y el casete de selección en regiones no codificantes que no afectan la regulación génica. A medida que los modelos animales se vuelven  
35 más complejos, pueden añadirse más modificaciones a las existentes, tales como mutaciones que causan enfermedades en humanos además de alelos humanizados. Los cambios adicionales pueden añadir aún más cicatrices y otro casete de selección a un locus de ratón ya altamente manipulado genéticamente, lo que aumenta la probabilidad de que la expresión se altere y el modelo de ratón no sea fiel a la enfermedad humana. Desde el punto de vista de la construcción, añadir un nuevo casete a un vector que ya contiene uno puede resultar complicado  
40 debido a la recombinación no deseada entre elementos de casete compartidos tales como promotores y señales poli(A), incluso si los dos casetes codifican selecciones diferentes. En consecuencia, se necesitan nuevos métodos para simplificar la generación de direccionamiento que porta múltiples cambios (tal como un alelo humanizado y una mutación de enfermedad superpuesta) y para minimizar las cicatrices incorporadas en un modelo animal final.

45 Resumen

Se proporcionan métodos de introducción sin cicatrices de una modificación genética dirigida en un vector de direccionamiento preexistente.

50 En un aspecto, algunos de tales métodos comprenden: (a) realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de modificación en una población de células bacterianas, en donde el casete de modificación comprende la modificación genética dirigida y comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de  
55 direccionamiento preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente, en donde el inserto de ácido nucleico comprende de 5' a 3': (i) una primera secuencia de repetición; (ii) un primer sitio diana para un primer agente nucleasa; (iii) un casete de selección; (iv) un segundo sitio diana para un segundo agente nucleasa; y (v) una segunda secuencia de repetición idéntica a la primera secuencia de repetición; (b) seleccionar células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección; (c) escindir el primer sitio diana en el vector de direccionamiento modificado con el primer agente nucleasa y escindir el segundo sitio diana en el vector de direccionamiento modificado con el segundo agente nucleasa para eliminar el casete de selección y exponer la primera secuencia de repetición y la segunda secuencia de repetición en el vector de direccionamiento modificado; y (d) ensamblar la primera secuencia de repetición expuesta con la segunda secuencia de repetición expuesta en una reacción de ensamblaje intramolecular  
60 *in vitro* para generar el vector de direccionamiento que comprenda la modificación genética dirigida sin cicatrices, en donde ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente  
65

nucleasa está presente y solo una única copia de la secuencia de repetición está presente en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices.

5 En algunos de tales métodos, la secuencia de repetición es idéntica a una secuencia en el vector de direccionamiento preexistente. En algunos de tales métodos, la modificación genética dirigida comprende una inserción, y la secuencia de repetición es idéntica al extremo 5' o al extremo 3' de la inserción.

10 En algunos de tales métodos, la secuencia de repetición tiene una longitud de al menos aproximadamente 20 nucleótidos. Opcionalmente, la secuencia de repetición tiene una longitud de entre aproximadamente 20 nucleótidos y aproximadamente 100 nucleótidos.

15 En algunos de tales métodos, el casete de modificación es un ácido nucleico lineal de doble cadena. En algunos de tales métodos, el casete de modificación tiene una longitud de aproximadamente 1 kb a aproximadamente 15 kb. En algunos de tales métodos, el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de al menos aproximadamente 35 nucleótidos. En algunos de tales métodos, el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 500 nucleótidos.

20 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa es un agente nucleasa de corte raro. En algunos de tales métodos, el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente. En algunos de tales métodos, el primer sitio diana es idéntico al segundo sitio diana y el primer agente nucleasa es idéntico al segundo agente nucleasa.

25 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa comprenden una enzima de restricción de corte raro. Opcionalmente, la enzima de restricción de corte raro es NotI, XmaIII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AbsI, CciNI, FspAI, MauBI, MreI, MssI, PaeAI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmlI, SrfI, Sse2321, Sse83871, LglI, PciSI, AarI, AjuI, AolI, BarI, PpiI o PspI.

30 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa es una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (Cas) y un ARN guía (ARNg), una nucleasa con dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN) o una meganucleasa manipulada genéticamente. Opcionalmente, el primer agente nucleasa y/o el segundo agente nucleasa, la proteína Cas y el ARNg, en donde la proteína Cas es Cas9, y en donde el ARNg comprende un ARN CRISPR (ARNcr) que direcciona y un ARN de CRISPR transactivador (ARNtracr).

35 En algunos de tales métodos, la modificación genética dirigida comprende una modificación en el brazo de homología 5' o el brazo de homología 3'. En algunos de tales métodos, la modificación genética dirigida comprende una modificación en el inserto de ácido nucleico. En algunos de tales métodos, la modificación genética dirigida comprende una mutación puntual, una eliminación, una inserción, un reemplazo o una combinación de los mismos.

40 En algunos de tales métodos, el casete de selección aporta resistencia a un antibiótico. Opcionalmente, el casete de selección aporta resistencia a ampicilina, cloranfenicol, tetraciclina, kanamicina, espectinomicina, estreptomycin, carbenicilina, bleomicina, eritromicina o polimixina B.

45 En algunos de tales métodos, el vector de direccionamiento preexistente es un gran vector de direccionamiento de una longitud de al menos aproximadamente 10 kb. Opcionalmente, el vector de direccionamiento preexistente tiene una longitud de al menos aproximadamente 100 kb.

50 En algunos de tales métodos, el vector de direccionamiento preexistente comprende un segundo casete de selección. Opcionalmente, el segundo casete de selección aporta resistencia a un antibiótico. Opcionalmente, el casete de selección en el casete de modificación y el segundo casete de selección en el vector de direccionamiento preexistente aportan cada uno resistencia a un antibiótico diferente. Opcionalmente, el segundo casete de selección permite la selección tanto en células bacterianas como de mamíferos.

55 En algunos de tales métodos, la etapa (c) se produce *in vitro*.

60 En algunos de tales métodos, la etapa (d) comprende: (i) poner en contacto el vector de direccionamiento modificado con una exonucleasa para exponer secuencias complementarias entre la primera secuencia de repetición y la segunda secuencia de repetición; (ii) hibridar las secuencias complementarias expuestas; (iii) extender los extremos 3' de las secuencias complementarias hibridadas; y (iv) ligar la secuencia complementaria hibridada. Opcionalmente, la etapa (d) comprende incubar el vector de direccionamiento modificado con una exonucleasa, una ADN polimerasa y una ADN ligasa.

65 Algunos de tales métodos comprenden además: (e) tratar el vector de direccionamiento con el primer agente nucleasa y el segundo agente nucleasa después del ensamblaje *in vitro* en la etapa (d) para verificar que ni el primer

sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes.

5 En otro aspecto, algunos de tales métodos comprenden: (a) realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de eliminación en una población de células bacterianas, en donde el casete de eliminación comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de direccionamiento preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente, en donde la secuencia diana 5' y la secuencia diana 3' flanquean una región del vector de direccionamiento preexistente en la que se va a introducir la modificación genética dirigida, y en donde el inserto de ácido nucleico comprende de 5' a 3': (i) un primer sitio diana para un primer agente nucleasa; (ii) un casete de selección; y (iii) un segundo sitio diana para un segundo agente nucleasa; (b) seleccionar células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección; (c) escindir el primer sitio diana en el vector de direccionamiento modificado con el primer agente nucleasa y escindir el segundo sitio diana en el vector de direccionamiento modificado con el segundo agente nucleasa para eliminar el casete de selección y exponer una secuencia final corriente arriba y una secuencia final corriente abajo en el vector de direccionamiento modificado; y (d) ensamblar en una reacción de ensamblaje *in vitro* el vector de direccionamiento escindido con un casete de modificación que comprende la modificación genética dirigida flanqueada por una secuencia final corriente arriba que se solapa con la secuencia final corriente arriba en el vector de direccionamiento modificado y una secuencia final corriente abajo que se solapa con la secuencia final corriente abajo en el vector de direccionamiento modificado para generar el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices, en donde ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices.

25 En algunos de tales métodos, el casete de eliminación tiene una longitud de aproximadamente 1 kb a aproximadamente 15 kb. En algunos de tales métodos, el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de al menos aproximadamente 35 nucleótidos. Opcionalmente, el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 500 nucleótidos. En algunos de tales métodos, el casete de eliminación es un ácido nucleico lineal de doble cadena.

30 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa es un agente nucleasa de corte raro. En algunos de tales métodos, el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente. En algunos de tales métodos, el primer sitio diana es idéntico al segundo sitio diana y el primer agente nucleasa es idéntico al segundo agente nucleasa.

35 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa comprenden una enzima de restricción de corte raro. Opcionalmente, la enzima de restricción de corte raro es NotI, XmaII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AbsI, CciNI, FspAI, MauBI, MreI, MssI, PaeI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmlI, SrfI, Sse2321, Sse83871, LglI, PciSI, AarI, AjuI, Alol, BarI, PpiI o PspI.

45 En algunos de tales métodos, el primer agente nucleasa y/o el segundo agente nucleasa es una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (Cas) y un ARN guía (ARNg), una nucleasa con dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN) o una meganucleasa manipulada genéticamente. Opcionalmente, el primer agente nucleasa y/o el segundo agente nucleasa, la proteína Cas y el ARNg, en donde la proteína Cas es Cas9, y en donde el ARNg comprende un ARN CRISPR (ARNcr) que direcciona y un ARN de CRISPR transactivador (ARNtracr).

50 En algunos de tales métodos, el casete de selección aporta resistencia a un antibiótico. Opcionalmente, el casete de selección aporta resistencia a ampicilina, cloranfenicol, tetraciclina, kanamicina, espectinomicina, estreptomycin, carbenicilina, bleomicina, eritromicina o polimixina B.

55 En algunos de tales métodos, el vector de direccionamiento preexistente es un gran vector de direccionamiento de una longitud de al menos 10 kb. Opcionalmente, el vector de direccionamiento preexistente tiene una longitud de al menos 100 kb.

60 En algunos de tales métodos, el vector de direccionamiento preexistente comprende un segundo casete de selección. Opcionalmente, el segundo casete de selección aporta resistencia a un antibiótico. Opcionalmente, el casete de selección en el casete de eliminación y el segundo casete de selección en el vector de direccionamiento preexistente aportan cada uno resistencia a un antibiótico diferente. Opcionalmente, el segundo casete de selección permite la selección tanto en células bacterianas como de mamíferos.

65 En algunos de tales métodos, la longitud de solapamiento entre la secuencia final corriente arriba en el casete de modificación y la secuencia final corriente arriba en el vector de direccionamiento modificado y/o la longitud de la solapamiento entre la secuencia final corriente abajo en el casete de modificación y la secuencia final corriente abajo en el vector de direccionamiento modificado tiene una longitud de al menos aproximadamente 20 nucleótidos. En

algunos de tales métodos, la longitud de solapamiento entre la secuencia final corriente arriba en el casete de modificación y la secuencia final corriente arriba en el vector de direccionamiento modificado y/o la longitud de la superposición entre la secuencia final corriente abajo en el casete de modificación y la secuencia final corriente abajo en el vector de direccionamiento modificado tiene una longitud de entre aproximadamente 20 y aproximadamente 100 nucleótidos.

En algunos de tales métodos, en donde la etapa (c) se produce *in vitro*.

En algunos de tales métodos, la etapa (d) comprende: (i) poner en contacto el vector de direccionamiento escindido y el casete de modificación con una exonucleasa para exponer secuencias complementarias entre las secuencias finales en el vector de direccionamiento modificado y las secuencias finales en el casete de modificación; (ii) hibridar las secuencias complementarias expuestas; (iii) extender los extremos 3' de las secuencias complementarias hibridadas; y (iv) ligar la secuencia complementaria hibridada. Opcionalmente, la etapa (d) comprende incubar el vector de direccionamiento escindido y el casete de modificación con una exonucleasa, una ADN polimerasa y una ADN ligasa.

En algunos de tales métodos, el casete de modificación es un ácido nucleico lineal de doble cadena. En algunos de tales métodos, el casete de modificación tiene una longitud de al menos aproximadamente 200 nucleótidos. En algunos de tales métodos, el casete de modificación tiene un tamaño que no puede sintetizarse o generarse directamente mediante la reacción en cadena de la polimerasa. En algunos de tales métodos, el casete de modificación tiene una longitud de al menos aproximadamente 10 kb.

En algunos de tales métodos, la modificación genética dirigida comprende una mutación puntual, una eliminación, una inserción, un reemplazo o una combinación de los mismos.

Algunos de tales métodos comprenden además: (e) tratar el vector de direccionamiento con el primer agente nucleasa y el segundo agente nucleasa después del ensamblaje *in vitro* en la etapa (d) para verificar que ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes.

Breve descripción de las figuras

La Figura 1 (no a escala) muestra un esquema de un método para la introducción sin cicatrices de una mutación puntual en un gran vector de direccionamiento a través de la recombinación homóloga bacteriana y el ensamblaje de Gibson intramolecular.

La Figura 2 (no a escala) muestra un esquema de un ácido nucleico sintetizado que se usará en el método mostrado en el esquema de la Figura 1.

La Figura 3 (no a escala) muestra un esquema de un método para la introducción sin cicatrices de una mutación puntual en un gran vector de direccionamiento a través de la recombinación homóloga bacteriana y el ensamblaje intermolecular de Gibson.

Las Figuras 4A-4B muestran una estrategia de direccionamiento tradicional mediante el uso de BAC modificados de ratón como vectores y tecnología de casete de eliminación automática, desde la construcción de vectores (Figura 4A) hasta la generación de ratones F1 (Figura 4B). La eliminación del casete a través de la recombinasa Cre expresada con protamina de ratón deja una cicatriz de 78 pb que contiene un solo loxP.

Definiciones

Los términos "proteína", "polipéptido" y "péptido", usados indistintamente en la presente descripción, incluyen formas poliméricas de aminoácidos de cualquier longitud, lo que incluye aminoácidos codificados y no codificados y aminoácidos modificados o derivatizados química o bioquímicamente. Los términos también incluyen polímeros que se han modificado, tales como polipéptidos que tienen estructuras peptídicas modificadas. El término "dominio" se refiere a cualquier parte de una proteína o polipéptido que tiene una función o estructura en particular.

Los términos "ácido nucleico" y "polinucleótido", usados indistintamente en la presente descripción, incluyen formas poliméricas de nucleótidos de cualquier longitud, lo que incluye ribonucleótidos, desoxirribonucleótidos o análogos o versiones modificadas de los mismos. Incluyen ADN o ARN de cadena única, de doble cadena y multicatenario, ADN genómico, ADNc, híbridos de ADN-ARN y polímeros que comprenden bases de purina, bases de pirimidina u otras bases nucleotídicas naturales, químicamente modificadas, bioquímicamente modificadas, no naturales o derivatizadas.

El término "vector de direccionamiento" se refiere a un ácido nucleico recombinante que puede introducirse mediante recombinación homóloga, ligadura mediada por unión de extremos no homólogos o cualquier otro medio de recombinación en una posición diana en el genoma de una célula.

El término "tipo silvestre" incluye entidades que tienen una estructura y/o actividad como la que se encuentra en un estado o contexto normal (por el contrario a mutante, enferma, alterada, etc.). Los genes y polipéptidos de tipo silvestre existen frecuentemente en múltiples formas diferentes (por ejemplo, alelos).

5 El término "secuencia endógena" se refiere a una secuencia de ácido nucleico que se produce naturalmente dentro de una célula o animal no humano. Por ejemplo, una secuencia endógena de *Rosa26* de un animal no humano se refiere a una secuencia nativa de *Rosa26* que se produce naturalmente en el locus *Rosa26* en el animal no humano.

10 Las moléculas o secuencias "exógenas" incluyen moléculas o secuencias que normalmente no están presentes en una célula en esa forma o localización (por ejemplo, locus genómico). La presencia normal incluye la presencia con relación a la etapa particular de desarrollo y las condiciones ambientales de la célula. Una molécula o secuencia exógena, por ejemplo, puede incluir una versión mutada de una secuencia endógena correspondiente dentro de la célula, tal como una versión humanizada de la secuencia endógena, o puede incluir una secuencia correspondiente a una secuencia endógena dentro de la célula pero en una forma diferente (es decir, no dentro de un cromosoma).  
15 Por el contrario, las moléculas o secuencias endógenas incluyen moléculas o secuencias que están presentes normalmente en esa forma y localización en una célula particular en una etapa de desarrollo particular en condiciones ambientales particulares.

20 El término "heterólogo", cuando se usa en el contexto de un ácido nucleico o una proteína, indica que el ácido nucleico o la proteína comprende al menos dos segmentos que no se producen naturalmente juntos en la misma molécula. Por ejemplo, el término "heterólogo", cuando se usa con referencia a segmentos de un ácido nucleico o segmentos de una proteína, indica que el ácido nucleico o la proteína comprende dos o más subsecuencias que no se encuentran en la naturaleza en la misma relación entre sí (por ejemplo, unidas). Como ejemplo, una región "heteróloga" de un vector de ácido nucleico es un segmento de ácido nucleico dentro o unido a otra molécula de  
25 ácido nucleico que no se encuentra asociada con la otra molécula en la naturaleza. Por ejemplo, una región heteróloga de un vector de ácido nucleico podría incluir una secuencia codificante flanqueada por secuencias que no se encuentran asociadas con la secuencia codificante en la naturaleza. Igualmente, una región "heteróloga" de una proteína es un segmento de aminoácidos dentro o unido a otra molécula peptídica que no se encuentra asociada con la otra molécula peptídica en la naturaleza (por ejemplo, una proteína de fusión o una proteína con una  
30 etiqueta). De manera similar, un ácido nucleico o una proteína pueden comprender un marcador heterólogo o una secuencia de localización o secreción heteróloga.

35 La "optimización de codones" se aprovecha de la degeneración de los codones, como lo demuestra la multiplicidad de combinaciones de codones de tres pares de bases que especifican un aminoácido, e incluye generalmente un proceso de modificación de una secuencia de ácido nucleico para potenciar la expresión en células huésped particulares mediante el reemplazo de al menos un codón de la secuencia nativa con un codón que se usa más frecuentemente o con mayor frecuencia en los genes de la célula huésped a la vez que se mantiene la secuencia de aminoácidos nativa. Por ejemplo, un ácido nucleico que codifica una proteína Cas9 puede modificarse para sustituir  
40 codones que tienen una mayor frecuencia de uso en una célula procarionota o eucariota determinada, lo que incluye una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster o cualquier otra célula huésped, en comparación con la secuencia de ácido nucleico de origen natural. Las tablas de uso de codones están fácilmente disponibles, por ejemplo, en la "Base de datos de uso de codones". Estas tablas pueden adaptarse de varias maneras. Ver Nakamura y otros (2000) *Nucleic Acids Res.* 28:292. También están disponibles algoritmos  
45 informáticos para la optimización de codones de una secuencia en particular para la expresión en un huésped particular (ver, por ejemplo, Gene Forge).

50 El término "locus" se refiere a una localización específica de un gen (o secuencia significativa), secuencia de ADN, secuencia codificante de polipéptido o posición en un cromosoma del genoma de un organismo. Por ejemplo, un locus *Rosa26* puede referirse a la localización específica de un gen *Rosa26*, una secuencia de ADN de *Rosa26* o una posición de *Rosa26* en un cromosoma del genoma de un organismo que se ha identificado en cuanto a dónde se encuentra tal secuencia. Un "locus *Rosa26*" puede comprender un elemento regulador de un gen *Rosa26*, que incluye, por ejemplo, un potenciador, un promotor, una región no traducida (UTR) 5' y/o 3', o una combinación de los  
55 mismos.

60 El término "gen" se refiere a una secuencia de ADN en un cromosoma que codifica un producto (por ejemplo, un producto de ARN y/o un producto polipeptídico) e incluye la región codificante interrumpida con intrones no codificantes y la secuencia localizada adyacente a la región codificante en los extremos 5' y 3' de manera que el gen corresponda al ARNm de longitud completa (lo que incluye las secuencias no traducidas 5' y 3'). El término "gen" también incluye otras secuencias no codificantes, lo que incluye secuencias reguladoras (por ejemplo, promotores, potenciadores y sitios de unión a factores de transcripción), señales de poliadenilación, sitios de entrada internos al ribosoma, silenciadores, secuencia aislante y regiones de unión a la matriz. Estas secuencias pueden estar cerca de la región codificante del gen (por ejemplo, dentro de 10 kb) o en sitios distantes, e influyen en el nivel o velocidad de transcripción y traducción del gen.  
65

Un "promotor" es una región reguladora de ADN que comprende generalmente una caja TATA capaz de dirigir a la ARN polimerasa II para iniciar la síntesis de ARN en el sitio de iniciación de la transcripción apropiado para una secuencia polinucleotídica particular. Un promotor puede comprender adicionalmente otras regiones que influyen en la velocidad de iniciación de la transcripción. Las secuencias promotoras descritas en la presente descripción modulan la transcripción de un polinucleótido unido operativamente. Un promotor puede ser activo en uno o más de los tipos de células descritos en la presente descripción (por ejemplo, una célula procariota o una célula eucariota (tal como una célula de mamífero), o una combinación de las mismas). Un promotor puede ser, por ejemplo, un promotor constitutivamente activo, un promotor condicional, un promotor inducible, un promotor temporalmente restringido (por ejemplo, un promotor regulado por el desarrollo) o un promotor espacialmente restringido (por ejemplo, un promotor específico de células o específico de tejidos). Pueden encontrarse ejemplos de promotores, por ejemplo, en el documento WO 2013/176772.

El "enlace operativo" o estar "unido operativamente" incluye la yuxtaposición de dos o más componentes (por ejemplo, un promotor y otro elemento de secuencia) de manera que ambos componentes funcionen normalmente y permitan la posibilidad de que al menos uno de los componentes pueda mediar en una función que se ejerce sobre al menos uno de los otros componentes. Por ejemplo, un promotor puede estar unido operativamente a una secuencia codificante si el promotor controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción. El enlace operativo puede incluir tales secuencias que son contiguas entre sí o que actúan *en trans* (por ejemplo, una secuencia reguladora puede actuar a distancia para controlar la transcripción de la secuencia codificante).

"Complementariedad" de ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos de nucleobase, forma enlaces de hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases complementarias en el ADN son típicamente A con T y C con G. En el ARN, son típicamente C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en el que cada base del dúplex se une a una base complementaria mediante el apareamiento de Watson-Crick. Complementario "sustancial" o "suficiente" significa que una secuencia en una cadena no es completa y/o perfectamente complementaria a una secuencia en una cadena opuesta, pero que se produce un enlace suficiente entre las bases de las dos cadenas para formar un complejo híbrido estable en un conjunto de condiciones de hibridación (por ejemplo, concentración de sal y temperatura). Tales condiciones pueden predecirse mediante el uso de las secuencias y los cálculos matemáticos estándar para predecir la  $T_m$  (temperatura de fusión) de las cadenas hibridadas, o mediante la determinación empírica de la  $T_m$  mediante el uso de métodos de rutina.  $T_m$  incluye la temperatura a la que una población de complejos de hibridación formados entre dos cadenas de ácido nucleico se desnaturaliza en un 50 % (es decir, una población de moléculas de ácido nucleico de doble cadena se disocia a la mitad en cadenas únicas). A una temperatura por debajo de la  $T_m$ , se favorece la formación de un complejo de hibridación, mientras que a una temperatura por encima de la  $T_m$ , se favorece la fusión o separación de las cadenas en el complejo de hibridación. La  $T_m$  puede estimarse para un ácido nucleico que tiene un contenido conocido de G+C en una solución acuosa de NaCl 1 M mediante el uso, por ejemplo, de  $T_m=81,5+0,41 (\% G+C)$ , aunque otros cálculos de  $T_m$  conocidos tienen en cuenta las características estructurales del ácido nucleico.

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles errores de apareamiento entre las bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y del grado de complementación, variables que son bien conocidas. Cuanto mayor sea el grado de complementación entre dos secuencias de nucleótidos, mayor será el valor de la temperatura de fusión ( $T_m$ ) para los híbridos de ácidos nucleicos que tienen esas secuencias. Para las hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo, complementariedad de más de 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos), la posición de los errores de apareamiento se vuelve importante (ver Sambrook y otros, más arriba, 11,7-11,8). Típicamente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Longitudes mínimas ilustrativas para un ácido nucleico hibridable incluyen al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 22 nucleótidos, al menos aproximadamente 25 nucleótidos y al menos aproximadamente 30 nucleótidos. Además, la temperatura y la concentración de sal de la solución de lavado pueden ajustarse según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

La secuencia del polinucleótido no necesita ser 100 % complementaria a la de su ácido nucleico diana para ser específicamente hibridable. Además, un polinucleótido puede hibridar sobre uno o más segmentos de manera que los segmentos intermedios o adyacentes no estén implicados en el evento de hibridación (por ejemplo, una estructura de bucle o una estructura de horquilla). Un polinucleótido (por ejemplo, ARNg) puede comprender al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 99 % o un 100 % de complementariedad de secuencia con una región diana dentro de la secuencia de ácido nucleico diana a la que se direccionan. Por ejemplo, un ARNg en el que 18 de 20 nucleótidos son complementarios a una región diana y, por lo tanto, se hibridarían específicamente, representaría un 90 % de complementariedad. En este ejemplo, los nucleótidos no complementarios restantes pueden agruparse o intercalarse con nucleótidos complementarios y no es necesario que sean contiguos entre sí o con nucleótidos complementarios.

El porcentaje de complementariedad entre tramos particulares de secuencias de ácido nucleico dentro de los ácidos nucleicos puede determinarse de forma rutinaria mediante el uso de programas BLAST (herramientas básicas de búsqueda de alineamiento local) y programas PowerBLAST (Altschul y otros (1990) *J. Mol. Biol.* 215:403-410; Zhang y Madden (1997) *Genoma Res.* 7:649-656) o mediante el uso del programa Gap (Wisconsin Sequence Analysis Package, Versión 8 de Unix, Genetics Computer Group, University Research Park, Madison Wis.), mediante el uso de la configuración predeterminada, que usa el algoritmo de Smith y Waterman (1981) *Adv. Appl. Math.* 2:482-489.

Los métodos y composiciones proporcionados en la presente descripción emplean una variedad de diferentes componentes. Algunos componentes a lo largo de la descripción pueden tener variantes y fragmentos activos. Tales componentes incluyen, por ejemplo, proteínas Cas, ARN CRISPR, ARNtracr y ARN guía. La actividad biológica de cada uno de estos componentes se describe en otra parte de la presente descripción. El término "funcional" se refiere a la capacidad innata de una proteína o ácido nucleico (o un fragmento o variante del mismo) para exhibir una actividad o función biológica. Tales actividades o funciones biológicas pueden incluir, por ejemplo, la capacidad de una proteína Cas para unirse a un ARN guía ya una secuencia de ADN diana. Las funciones biológicas de los fragmentos o variantes funcionales pueden ser las mismas o, de hecho, pueden cambiar (por ejemplo, con respecto a su especificidad, selectividad o eficacia) en comparación con el original, pero conservando la función biológica básica.

El término "variante" se refiere a una secuencia de nucleótidos que difiere de la secuencia más prevalente en una población (por ejemplo, en un nucleótido) o una secuencia de proteína diferente de la secuencia más prevalente en una población (por ejemplo, en un aminoácido).

El término "fragmento" cuando se refiere a una proteína significa una proteína que es más corta o tiene menos aminoácidos que la proteína completa. El término "fragmento" cuando se refiere a un ácido nucleico significa un ácido nucleico que es más corto o tiene menos nucleótidos que el ácido nucleico de longitud completa. Un fragmento puede ser, por ejemplo, un fragmento N-terminal (es decir, eliminación de una porción del extremo C-terminal de la proteína), un fragmento C-terminal (es decir, eliminación de una porción del extremo N-terminal de la proteína), o un fragmento interno.

"Identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias polipeptídicas hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean para una máxima correspondencia en una ventana de comparación especificada. Cuando se usa el porcentaje de identidad de secuencia en referencia a las proteínas, las posiciones de los residuos que no son idénticas frecuentemente difieren por sustituciones conservadoras de aminoácidos, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y, por lo tanto, no cambia las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservadoras, el por ciento de identidad de secuencia puede ajustarse al alza para corregir la naturaleza conservadora de la sustitución. Se dice que las secuencias que difieren por dichas sustituciones conservadoras tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos. Típicamente, esto implica calificar una sustitución conservadora como un error de coincidencia parcial en lugar de total, lo que aumenta de esta manera el porcentaje de identidad de secuencia. De esta forma, por ejemplo, cuando un aminoácido idéntico recibe una puntuación de 1 y una sustitución no conservadora recibe una puntuación de cero, una sustitución conservadora recibe una puntuación entre cero y 1. La calificación de las sustituciones conservadoras se calcula, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

El "porcentaje de identidad de secuencia" incluye el valor determinado mediante la comparación de dos secuencias alineadas de forma óptima (el mayor número de residuos perfectamente coincidentes) en una ventana de comparación, en donde la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) para un alineamiento óptimo de las dos secuencias. El porcentaje se calcula al determinar el número de posiciones en las que aparecen residuos de aminoácidos o bases de ácidos nucleicos idénticas en ambas secuencias para producir el número de posiciones coincidentes, dividir el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicar el resultado por 100 para producir el porcentaje de identidad de secuencia. A menos que se especifique lo contrario (por ejemplo, la secuencia más corta incluye una secuencia heteróloga unida), la ventana de comparación es la longitud total de la más corta de las dos secuencias que se comparan.

A menos que se indique lo contrario, los valores de identidad/similitud de secuencia incluyen el valor obtenido mediante el uso de la versión 10 de GAP con los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos mediante el uso de un peso de hueco de 50 y un peso de longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos mediante el uso de un peso de hueco de 8 y un peso de longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente del mismo. "Programa equivalente" incluye cualquier programa de comparación de secuencias que, para cualquiera de dos secuencias en cuestión, genera un alineamiento que tiene coincidencias idénticas de residuos de

aminoácidos o nucleótidos y un por ciento de identidad de secuencia idéntico en comparación con el alineamiento correspondiente generado mediante la versión 10 de GAP.

El término "sustitución conservadora de aminoácidos" se refiere a la sustitución de un aminoácido que está presente normalmente en la secuencia con un aminoácido diferente de similar tamaño, carga o polaridad. Los ejemplos de sustituciones conservadoras incluyen la sustitución de un residuo no polar (hidrófobo) tal como isoleucina, valina o leucina por otro residuo no polar. Igualmente, los ejemplos de sustituciones conservadoras incluyen la sustitución de un residuo polar (hidrófilo) por otro, tal como entre arginina y lisina, entre glutamina y asparagina, o entre glicina y serina. Adicionalmente, la sustitución de un residuo básico tal como lisina, arginina o histidina por otro, o la sustitución de un residuo ácido tal como ácido aspártico o ácido glutámico por otro residuo ácido, son ejemplos adicionales de sustituciones conservadoras. Los ejemplos de sustituciones no conservadoras incluyen la sustitución de un residuo de aminoácido no polar (hidrófobo) tal como isoleucina, valina, leucina, alanina o metionina por un residuo polar (hidrófilo) tal como cisteína, glutamina, ácido glutámico o lisina y/o un residuo polar por un residuo no polar. Las categorizaciones típicas de aminoácidos se resumen en la Tabla 1 a continuación.

**Tabla 1. Categorizaciones de aminoácidos.**

Alanina	Ala	A	No polar	Neutro	1,8
Arginina	Arg	R	Polar	Positivo	-4,5
Asparagina	Asn	N	Polar	Neutro	-3,5
Ácido aspártico	Asp	D	Polar	Negativo	-3,5
Cisteína	Cys	C	No polar	Neutro	2,5
Ácido glutámico	Glu	E	Polar	Negativo	-3,5
Glutamina	Gln	Q	Polar	Neutro	-3,5
Glicina	Gly	G	No polar	Neutro	-0,4
Histidina	His	H	Polar	Positivo	-3,2
Isoleucina	Ile	I	No polar	Neutro	4,5
Leucina	Leu	L	No polar	Neutro	3,8
Lisina	Lys	K	Polar	Positivo	-3,9
Metionina	Met	M	No polar	Neutro	1,9
Fenilalanina	Phe	F	No polar	Neutro	2,8
Prolina	Pro	P	No polar	Neutro	-1,6
Serina	Ser	S	Polar	Neutro	-0,8
Treonina	Thr	T	Polar	Neutro	-0,7
Triptófano	Trp	W	No polar	Neutro	-0,9
Tirosina	Tyr	Y	Polar	Neutro	-1,3
Valina	Val	V	No polar	Neutro	4,2

Una secuencia "homóloga" (por ejemplo, secuencia de ácido nucleico) incluye una secuencia que es idéntica o sustancialmente similar a una secuencia de referencia conocida, de manera que es, por ejemplo, al menos 50 %, al menos 55 %, al menos 60 %, al menos 65 %, al menos 70 %, al menos 75 %, al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a la secuencia de referencia conocida. Las secuencias homólogas pueden incluir, por ejemplo, secuencias ortólogas y secuencias parálogas. Los genes homólogos, por ejemplo, descienden típicamente de una secuencia de ADN ancestral común, ya sea mediante un evento de especiación (genes ortólogos) o un evento de duplicación genética (genes parálogos). Los genes "ortólogos" incluyen genes en diferentes especies que evolucionaron a partir de un gen ancestral común por especiación. Los ortólogos conservan típicamente la misma función en el curso de la evolución. Los genes "parálogos" incluyen genes relacionados por duplicación dentro de un genoma. Los parálogos pueden desarrollar nuevas funciones en el curso de la evolución.

El término "*in vitro*" incluye ambientes artificiales y procesos o reacciones que se producen dentro de un ambiente artificial (por ejemplo, un tubo de ensayo). El término "*in vivo*" incluye ambientes naturales (por ejemplo, una célula u organismo o cuerpo) y los procesos o reacciones que se producen dentro de un ambiente natural. El término "*ex vivo*" incluye células que se han extraído del cuerpo de un individuo y procesos o reacciones que se producen dentro de tales células.

La reparación en respuesta a rupturas de doble cadena (DSB) se produce principalmente mediante dos vías de reparación de ADN conservadas: recombinación homóloga (HR) y unión de extremos no homólogos (NHEJ). Ver Kasperek y Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897. Igualmente, la reparación de un ácido nucleico diana mediada por un ácido nucleico donante exógeno puede incluir cualquier proceso de intercambio de información genética entre los dos polinucleótidos.

El término "recombinación" incluye cualquier proceso de intercambio de información genética entre dos polinucleótidos y puede producirse por cualquier mecanismo. La recombinación puede producirse a través de la reparación dirigida por homología (HDR) o la recombinación homóloga (HR). La HDR o la HR incluyen una forma de reparación de ácido nucleico que puede requerir homología de secuencia de nucleótidos, usa una molécula "donante" como plantilla para la reparación de una molécula "diana" (es decir, la que experimentó la ruptura de doble

cadena) y conduce a la transferencia de información genética del donante a la diana. Sin pretender limitarse por ninguna teoría en particular, dicha transferencia puede implicar la corrección del error de coincidencia del ADN heterodúplex que se forma entre la diana rota y el donante, y/o la hibridación de la cadena dependiente de la síntesis, en la que el donante se usa para resintetizar la información genética que formará parte de la diana y/o procesos relacionados. En algunos casos, el polinucleótido donante, una porción del polinucleótido donante, una copia del polinucleótido donante o una porción de una copia del polinucleótido donante se integra en el ADN diana. Ver Wang y otros (2013) Cell 153:910-918; Mandalos y otros (2012) PLOS ONE 7:e45768:1-9; y Wang y otros (2013) Nat Biotechnol. 31:530-532.

Las composiciones o métodos "que comprenden" o "que incluyen" uno o más elementos enumerados pueden incluir otros elementos no enumerados específicamente. Por ejemplo, una composición que "comprende" o "incluye" una proteína puede contener la proteína sola o en combinación con otros ingredientes. La frase de transición "que consiste esencialmente en" significa que el alcance de una reivindicación debe interpretarse como que abarca los elementos enumerados especificados en la reivindicación y aquellos que no afectan materialmente la(s) característica(s) básica(s) y novedosa(s) de la invención reivindicada. Por tanto, el término "que consiste esencialmente en" cuando se usa en una reivindicación de esta invención no pretende ser interpretado como equivalente a "que comprende".

"Opcional" u "opcionalmente" significa que el evento o circunstancia descrito subsecuentemente puede producirse o no y que la descripción incluye casos en los que se produce el evento o circunstancia y casos en los que no se produce.

La designación de un intervalo de valores incluye todos los números enteros dentro o que definen el intervalo, y todos los subintervalos definidos por números enteros dentro del intervalo.

A menos que sea evidente de otra manera por el contexto, el término "aproximadamente" abarca valores dentro de un margen estándar de error de medición (por ejemplo, SEM) de un valor indicado.

El término "y/o" se refiere y abarca todas y cada una de las combinaciones posibles de uno o más de los elementos enumerados asociados, así como también la ausencia de combinaciones cuando se interpreta en la alternativa ("o").

El término "o" se refiere a cualquier miembro de una lista en particular y también incluye cualquier combinación de miembros de esa lista.

Las formas singulares de los artículos "un", "una" y "el/la" incluye referencias en plural a menos que el contexto lo indique claramente de cualquier otra manera. Por ejemplo, el término "una proteína" o "al menos una proteína" puede incluir una pluralidad de proteínas, lo que incluye sus mezclas.

Estadísticamente significativo significa  $p \leq 0,05$ .

Descripción detallada

### *I. Visión general*

En la presente descripción se proporcionan métodos para introducir una modificación genética dirigida sin cicatrices en un vector de direccionamiento preexistente. Los métodos pueden usar combinaciones de recombinación homóloga bacteriana (BHR) y métodos de ensamblaje *in vitro* (ya sea intramolecular o intermolecular) para introducir tales modificaciones genéticas dirigidas en un vector de direccionamiento sin cicatrices. El término sin cicatrices se refiere al hecho de que las reacciones no introducen cambios ni secuencias no deseadas en el ADN ensamblado. La secuencia combinada corresponderá a la secuencia exacta deseada sin cambios ni artefactos introducidos por los procedimientos de ensamblaje *in vitro* o BHR.

Uno de los enfoques más efectivos para determinar la función de los genes implica las mutaciones genéticas deliberadas por manipulación genética en células madre embrionarias (ES) de ratón (u otras células ES de animales no humanos), y a continuación generar ratones (u otros animales no humanos) que albergan los genéticos correspondientes. As dos etapas limitantes son la generación de vectores de direccionamiento a genes y la posterior selección de clones de células ES raros en los que el vector de direccionamiento ha alterado correctamente el gen. Para producir una alteración genética deseada en las células ES, primero se debe introducir la alteración en un vector de direccionamiento que se usa posteriormente para reemplazar el gen nativo en las células ES mediante recombinación homóloga.

La construcción de ADN sin cicatrices es de particular importancia cuando se crean líneas de animales transgénicos, ya que las cicatrices producidas por sitios de restricción u otras manipulaciones pueden tener un impacto negativo en la expresión génica si se producen en una región importante para la regulación. El direccionamiento al genoma de los mamíferos frecuentemente requiere la construcción de grandes vectores de direccionamiento con brazos de ADN largos para dirigir la recombinación homóloga, así como también casetes de resistencia a antibióticos para la

selección de clones de células madre embrionarias. Los clones dirigidos correctamente frecuentemente contienen múltiples cicatrices necesarias para la construcción del vector y el propio casete de resistencia. Incluso con la tecnología de casete con eliminación automática, frecuentemente no es posible evitar dejar "cicatrices" de secuencias exógenas en loci modificados. *Ver, por ejemplo, las Figuras 4A-4B.* Tales cicatrices pueden afectar la expresión fiel del locus dirigido o incluso la expresión de genes vecinos. A medida que los modelos animales se vuelven más complejos, pueden añadirse más modificaciones a las existentes, tales como mutaciones que causan enfermedades humanas en alelos humanizados. Los cambios adicionales pueden añadir aún más cicatrices y otro casete de selección a un locus de ratón ya altamente manipulado genéticamente, lo que aumenta la probabilidad de que la expresión se altere y el modelo de ratón no sea fiel. Además, añadir un nuevo casete a un vector que ya contiene uno puede complicarse debido a la recombinación no deseada entre elementos de casete compartidos tales como promotores y señales poli(A), incluso si los dos casetes codifican selecciones diferentes. Sin embargo, tales casetes de selección son importantes para que no se desperdicien tiempo ni recursos en el tamizaje de miles de clones de células ES para una modificación deseada.

Alternativamente, usar el vector de direccionamiento inicial para crear y seleccionar células ES modificadas que comprenden la modificación del vector de direccionamiento inicial y a continuación volver a dirigir esas células con un segundo vector de direccionamiento (por ejemplo, ssODN) para realizar una segunda modificación en el locus ya dirigido requiere mucho tiempo y el redireccionamiento (por ejemplo, con ssODN) puede dar lugar a modificaciones no deseadas tales como inserciones no deseadas, eliminaciones no deseadas, mutaciones puntuales no deseadas o falta de direccionamiento junto con una inserción transgénica en otra parte del genoma.

Los métodos descritos en la presente descripción proporcionan métodos eficientes y sin cicatrices para realizar modificaciones en vectores de direccionamiento preexistentes en la etapa de preparación del vector de direccionamiento en lugar de tener que crear y tamizar células ES que comprenden el vector de direccionamiento preexistente inicial y a continuación volver a dirigir esas células para hacer un segunda modificación al locus ya dirigido.

II. Introducción sin cicatrices de una modificación dirigida en un vector de direccionamiento mediante recombinación homóloga bacteriana y ensamblaje intramolecular *in vitro*

Algunos métodos descritos en la presente descripción para la introducción sin cicatrices de una modificación genética dirigida en un vector de direccionamiento preexistente aprovechan los métodos de ensamblaje *in vitro* para el ensamblaje intramolecular. Como ejemplo, tales métodos pueden comprender realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de modificación en una población de células bacterianas. El casete de modificación puede comprender un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de direccionamiento preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente. El inserto de ácido nucleico puede comprender un casete de selección flanqueado por sitios diana para uno o más agentes nucleasa (por ejemplo, agentes nucleasa de corte raro) y secuencias de repetición. Por ejemplo, el inserto de ácido nucleico puede comprender de 5' a 3': (1) una primera secuencia de repetición; (2) un primer sitio diana para un primer agente nucleasa; (3) un casete de selección; (4) un segundo sitio diana para un segundo agente nucleasa; y (5) una segunda secuencia de repetición.

El vector de direccionamiento preexistente puede ser cualquier tipo de vector de direccionamiento de cualquier tamaño. En un ejemplo específico, el vector de direccionamiento preexistente es un vector de direccionamiento grande (LTVEC) que tiene una longitud de al menos aproximadamente 10 kb. En otro ejemplo, tiene una longitud de al menos aproximadamente 100 kb. Los vectores de direccionamiento y los vectores de direccionamiento grandes se analizan con más detalle en otra parte de la presente descripción.

El casete de modificación puede ser un ácido nucleico lineal o un ácido nucleico circular, puede ser un ácido nucleico de cadena única o un ácido nucleico de doble cadena y puede comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). En un ejemplo específico, el casete de modificación es un ADN lineal de doble cadena.

Los brazos de homología en el casete de modificación se denominan en la presente descripción brazos de homología 5' y 3' (es decir, corriente arriba y corriente abajo). Esta terminología se refiere a la posición relativa de los brazos de homología con relación al inserto de ácido nucleico dentro de casete de modificación. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus de direccionamiento preexistente a modificar, que se denominan en la presente descripción "secuencia diana 5'" y "secuencia diana 3'", respectivamente.

Un brazo de homología y una secuencia diana "corresponden" o "se corresponden" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga (por ejemplo, recombinación homóloga bacteriana). El término "homología" incluye secuencias de ADN que son idénticas o comparten identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre una secuencia diana dada y el brazo de homología correspondiente que se encuentra en la plantilla de reparación exógena puede ser cualquier grado de identidad de secuencia que permita que se produzca la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo

de homología de la plantilla de reparación exógena (o un fragmento de la misma) y la secuencia diana (o un fragmento de la misma) puede ser al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de manera que las secuencias experimentan una recombinación homóloga. Además, una región de homología correspondiente entre el brazo de homología y la secuencia diana correspondiente puede tener cualquier longitud que sea suficiente para promover la recombinación homóloga. Por ejemplo, los brazos de homología pueden ser de cualquier tamaño adecuado para la recombinación homóloga bacteriana. Por ejemplo, los brazos de homología pueden tener al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 100 nucleótidos. Por ejemplo, los brazos de homología pueden tener entre aproximadamente 35 nucleótidos y 500 nucleótidos, entre aproximadamente 75 nucleótidos y aproximadamente 500 nucleótidos, o entre aproximadamente 50 nucleótidos y aproximadamente 200 nucleótidos (por ejemplo, aproximadamente 100 nucleótidos). Como otro ejemplo, los brazos de homología pueden tener una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 2,5 kb, una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 1,5 kb, o tienen una longitud de entre aproximadamente 35 y aproximadamente 500 nucleótidos. Por ejemplo, un brazo de homología dado (o cada uno de los brazos de homología) y/o la secuencia diana correspondiente puede comprender regiones de homología correspondientes que tienen una longitud entre aproximadamente 35 y aproximadamente 40, aproximadamente 40 a aproximadamente 50, aproximadamente 50 a aproximadamente 60, aproximadamente 60 y aproximadamente 70, aproximadamente 70 y aproximadamente 80, aproximadamente 80 y aproximadamente 90, aproximadamente 90 y aproximadamente 100, aproximadamente 100 y aproximadamente 150, aproximadamente 150 y aproximadamente 200, aproximadamente 200 y aproximadamente 250, aproximadamente 250 y aproximadamente 300, aproximadamente 300 y aproximadamente 350, aproximadamente 350 y aproximadamente 400, aproximadamente 400 y aproximadamente 450, o aproximadamente 450 y aproximadamente 500 nucleótidos, de manera que los brazos de homología tengan suficiente homología para sufrir una recombinación homóloga con las secuencias diana correspondientes dentro del ácido nucleico diana. Alternativamente, un brazo de homología dado (o cada brazo de homología) y/o la secuencia diana correspondiente pueden comprender regiones de homología correspondientes que tienen una longitud de entre aproximadamente 0,5 kb y aproximadamente 1 kb, aproximadamente 1 kb y aproximadamente 1,5 kb, aproximadamente 1,5 kb y aproximadamente 2 kb, o aproximadamente 2 kb a aproximadamente 2,5 kb. Por ejemplo, los brazos de homología pueden tener cada uno una longitud de aproximadamente 100 nucleótidos. Los brazos de homología pueden ser simétricos (cada uno aproximadamente del mismo tamaño en longitud), o pueden ser asimétricos (uno más largo que el otro).

El casete de modificación puede tener cualquier longitud. Por ejemplo, un casete de modificación puede ser de aproximadamente 10 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. En un ejemplo, un casete de modificación puede tener una longitud de al menos aproximadamente 100 kb o 100 kb. Un casete de modificación también puede ser de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 100 kb a aproximadamente 125 kb, de aproximadamente 300 kb a aproximadamente 325 kb, de aproximadamente 325 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 375 kb, de aproximadamente 375 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 425 kb, de aproximadamente 425 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 475 kb, o de aproximadamente 475 kb a aproximadamente 500 kb. Alternativamente, un casete de modificación puede tener al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más. En un ejemplo, el casete de modificación tiene una longitud de entre aproximadamente 1 kb y aproximadamente 15 kb o una longitud de entre aproximadamente 1 kb y aproximadamente 10 kb (por ejemplo, aproximadamente 1,2 kb, aproximadamente 5 kb, aproximadamente 8 kb o aproximadamente 15 kb).

El casete de modificación puede comprender la modificación genética dirigida. Por ejemplo, la modificación genética dirigida puede estar en el brazo de homología 5' o en el brazo de homología 3' (por ejemplo, una pequeña modificación tal como una mutación puntual o una pequeña eliminación, inserción o reemplazo que no afectará negativamente la capacidad del brazo de homología para recombinarse con la secuencia diana). Alternativamente, la modificación genética dirigida puede estar en el inserto de ácido nucleico (por ejemplo, cuando la modificación genética dirigida es una inserción o un reemplazo). Si la única modificación genética dirigida es una eliminación, entonces el brazo de homología 5' y el brazo de homología 3' pueden diseñarse para apuntar a las secuencias diana 5' y 3', respectivamente, que flanquean la secuencia dirigida para la eliminación en el vector de direccionamiento

preexistente. Como ejemplo, la modificación genética dirigida puede estar en la primera secuencia de repetición y/o en la segunda secuencia de repetición en el inserto de ácido nucleico. Los tipos de posibles modificaciones genéticas dirigidas se describen con más detalle en otra parte de la presente descripción. Algunos ejemplos incluyen mutaciones puntuales, eliminaciones, inserciones, reemplazos o combinaciones de los mismos.

Las secuencias de repetición primera y segunda en el casete de modificación pueden ser idénticas entre sí. La secuencia de repetición puede ser idéntica a una secuencia en el vector de direccionamiento preexistente. Alternativamente, en el caso de que la modificación genética dirigida comprenda una inserción (por ejemplo, una inserción sola o una inserción en combinación con una eliminación (es decir, reemplazo)), la secuencia de repetición puede ser idéntica al extremo 5' o al extremo 3' de la inserción.

La secuencia de repetición puede tener cualquier tamaño adecuado para el ensamblaje posterior entre las secuencias de repetición primera y segunda en una reacción de ensamblaje *in vitro*. Como ejemplo, la secuencia de repetición puede comprender al menos aproximadamente 20 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 40 nucleótidos o al menos aproximadamente 50 nucleótidos. Como otro ejemplo, la secuencia de repetición puede tener una longitud de entre aproximadamente 20 nucleótidos y aproximadamente 100 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 90 nucleótidos, entre aproximadamente 0 nucleótidos y aproximadamente 80 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 70 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 60 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 50 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 40 nucleótidos, entre aproximadamente 30 nucleótidos y aproximadamente 60 nucleótidos, o entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos. En un ejemplo específico, la secuencia de repetición puede tener una longitud de entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos (por ejemplo, aproximadamente 40 nucleótidos o aproximadamente 50 nucleótidos).

Después de la recombinación homóloga bacteriana, pueden seleccionarse células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección (y que comprende la modificación genética dirigida). Los ejemplos de casetes de selección y métodos de selección se describen con más detalle en otra parte de la presente descripción. En un ejemplo específico, el casete de selección aporta resistencia a un antibiótico. Por ejemplo, puede aportar resistencia a cualquiera de ampicilina, cloranfenicol, tetraciclina, kanamicina, espectinomicina, estreptomycinina, carbenicilina, bleomicina, eritromicina o polimixina B. En algunos métodos, el vector de direccionamiento preexistente también comprende un segundo casete de selección. El segundo casete de selección también puede aportar, por ejemplo, resistencia a un antibiótico. El casete de selección en el casete de modificación y el segundo casete de selección en el vector de direccionamiento preexistente pueden aportar cada uno resistencia a un antibiótico diferente. Por ejemplo, el casete de selección en el casete de modificación puede aportar resistencia a un primer antibiótico, y el segundo casete de selección en el vector de direccionamiento preexistente puede aportar resistencia a un segundo antibiótico diferente. En algunos métodos, el segundo casete de selección puede permitir la selección tanto en células bacterianas como en células eucarióticas o de mamíferos.

Después de la selección, el primer sitio diana en el vector de direccionamiento modificado puede dividirse con el primer agente nucleasa y el segundo sitio diana en el vector de direccionamiento modificado puede dividirse con el segundo agente nucleasa para eliminar el casete de selección y exponer la primera secuencia de repetición y la segunda secuencia de repetición en el vector de direccionamiento modificado. Por ejemplo, esta etapa puede hacerse *in vitro*. Como ejemplo, el ADN puede aislarse de las células bacterianas después de la recombinación homóloga bacteriana y selección, después de lo cual el primer sitio diana en el vector de direccionamiento modificado puede escindirse con el primer agente nucleasa *in vitro*, y el segundo sitio diana en el vector de direccionamiento modificado puede escindirse con el segundo agente nucleasa *in vitro* para eliminar el casete de selección y exponer la primera secuencia de repetición y la segunda secuencia de repetición en el vector de direccionamiento modificado.

El primer agente nucleasa y/o el segundo agente nucleasa pueden ser un agente nucleasa de corte raro como se describe en otra parte de la presente descripción. Por ejemplo, en algunos métodos, el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente. Los sitios diana primero y segundo pueden ser diferentes, o el primer sitio diana puede ser idéntico al segundo sitio diana, y el primer agente nucleasa puede ser idéntico al segundo agente nucleasa. El primer agente nucleasa y/o el segundo agente nucleasa pueden crear un extremo romo, una protuberancia 5' o una protuberancia 3'. En un ejemplo, el primer agente nucleasa y/o el segundo agente nucleasa crean una protuberancia 3'.

En un ejemplo específico, el primer agente nucleasa y/o el segundo agente nucleasa es una enzima de restricción o una enzima de restricción de corte raro. Los ejemplos de enzimas de restricción de corte raro se describen en otra parte de la presente descripción, pero pueden incluir, por ejemplo, NotI, XmaII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AbsI, CciNI, FspAI, MauBI, MreI, MssI, PaeAI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmiI, SrfI, Sse2321, Sse83871, LguI, PciSI, AarI, AjuI, AolI, BarI, PpiI y PsrI.

En otro ejemplo específico, el primer agente nucleasa y/o el segundo agente nucleasa pueden ser un agente nucleasa manipulado genéticamente. Por ejemplo, el agente nucleasa puede ser una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (CAS) y un ARN guía (ARNg) (por ejemplo, Cas9 y un ARNg que comprende un ARN CRISPR (ARNcr) y un ARN CRISPR transactivador (ARNtracr)), una nucleasa con dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN) o una meganucleasa manipulada genéticamente

Después de la escisión/digestión, la primera secuencia de repetición expuesta puede ensamblarse con la segunda secuencia de repetición expuesta en una reacción de ensamblaje intramolecular *in vitro* para generar el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices. Por ejemplo, en algunos de tales métodos, ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices (es decir, después del ensamblaje *in vitro*). Del mismo modo, en algunos de tales métodos, sólo está presente una única copia de la secuencia de repetición en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices (es decir, después del ensamblaje *in vitro*).

Puede usarse cualquier método de ensamblaje *in vitro* adecuado. En un ejemplo específico, la etapa de ensamblaje *in vitro* puede comprender la incubación del vector de direccionamiento modificado con una exonucleasa, una ADN polimerasa y una ADN ligasa. Por ejemplo, el método de ensamblaje *in vitro* puede comprender poner en contacto el vector de direccionamiento modificado con una exonucleasa para exponer secuencias complementarias entre la primera secuencia de repetición y la segunda secuencia de repetición, hibridar las secuencias complementarias expuestas, extender los extremos 3' de las secuencias complementarias hibridadas, y ligar las secuencias complementarias hibridadas. Los ejemplos de métodos de ensamblaje *in vitro* se analizan con más detalle en otra parte de la presente descripción.

En algunos métodos, para reducir el fondo, el vector producido por el ensamblaje *in vitro* puede tratarse con el primer agente nucleasa y/o el segundo agente nucleasa para reducir el fondo (por ejemplo, mediante la escisión de cualquiera de los vectores de direccionamiento que no se ensamblan exitosamente y por lo tanto aún contienen el sitio diana para el primer agente nucleasa o el segundo agente nucleasa). Tal etapa puede ayudar a verificar que no estén presentes ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa.

### *III. Introducción sin cicatrices de una modificación dirigida en un vector de direccionamiento a través de la recombinación homóloga bacteriana y el ensamblaje in vitro intermolecular*

Otros métodos descritos en la presente descripción para la introducción sin cicatrices de una modificación genética dirigida en un vector de direccionamiento preexistente aprovechan los métodos de ensamblaje *in vitro* para el ensamblaje intermolecular. Como ejemplo, tales métodos pueden comprender realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de eliminación en una población de células bacterianas. El casete de eliminación puede comprender un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de diana preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente. El clan de secuencias diana 5' y secuencias diana 3' flanquean una región del vector de direccionamiento preexistente en el que se va a introducir la modificación genética dirigida. El inserto de ácido nucleico puede comprender un casete de selección flanqueado por sitios diana para uno o más agentes de nucleasa (por ejemplo, agentes de nucleasa de corte raro). Por ejemplo, el inserto de ácido nucleico puede comprender de 5' a 3': (1) un primer sitio diana para un primer agente nucleasa; (2) un casete de selección; y (3) un segundo sitio diana para un segundo agente nucleasa.

El vector de direccionamiento preexistente puede ser cualquier tipo de vector de direccionamiento de cualquier tamaño. En un ejemplo específico, el vector de direccionamiento preexistente es un vector de direccionamiento grande (LTVEC) que tiene una longitud de al menos aproximadamente 10 kb. En otro ejemplo, tiene una longitud de al menos aproximadamente 100 kb. Los vectores de direccionamiento se analizan con más detalle en otra parte de la presente descripción.

El casete de eliminación puede ser un ácido nucleico lineal o un ácido nucleico circular, puede ser un ácido nucleico de cadena única o un ácido nucleico de doble cadena y puede comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). En un ejemplo específico, el casete de modificación es un ADN lineal de doble cadena.

Los brazos de homología en el casete de eliminación se denominan en la presente descripción brazos de homología 5' y 3' (es decir, corriente arriba y corriente abajo). Esta terminología se refiere a la posición relativa de los brazos de homología con relación al inserto de ácido nucleico dentro de casete de eliminación. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus de direccionamiento preexistente a modificar, que se denominan en la presente descripción "secuencia diana 5'" y "secuencia diana 3'", respectivamente.

Un brazo de homología y una secuencia diana "corresponden" o "se corresponden" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de

recombinación homóloga (por ejemplo, recombinación homóloga bacteriana). El término "homología" incluye secuencias de ADN que son idénticas o comparten identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre una secuencia diana dada y el brazo de homología correspondiente que se encuentra en la plantilla de reparación exógena puede ser cualquier grado de identidad de secuencia que permita que se produzca la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología de la plantilla de reparación exógena (o un fragmento de la misma) y la secuencia diana (o un fragmento de la misma) puede ser al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de manera que las secuencias experimentan una recombinación homóloga. Además, una región de homología correspondiente entre el brazo de homología y la secuencia diana correspondiente puede tener cualquier longitud que sea suficiente para promover la recombinación homóloga. Por ejemplo, los brazos de homología pueden ser de cualquier tamaño adecuado para la recombinación homóloga bacteriana. Por ejemplo, los brazos de homología pueden tener al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 100 nucleótidos. Por ejemplo, los brazos de homología pueden tener entre aproximadamente 35 nucleótidos y 500 nucleótidos, entre aproximadamente 75 nucleótidos y aproximadamente 500 nucleótidos, o entre aproximadamente 50 nucleótidos y aproximadamente 200 nucleótidos (por ejemplo, aproximadamente 100 nucleótidos). Como otro ejemplo, los brazos de homología pueden tener una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 2,5 kb, una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 1,5 kb, o tienen una longitud de entre aproximadamente 35 y aproximadamente 500 nucleótidos. Por ejemplo, un brazo de homología dado (o cada uno de los brazos de homología) y/o la secuencia diana correspondiente puede comprender regiones de homología correspondientes que tienen una longitud entre aproximadamente 35 y aproximadamente 40, aproximadamente 40 a aproximadamente 50, aproximadamente 50 a aproximadamente 60, aproximadamente 60 y aproximadamente 70, aproximadamente 70 y aproximadamente 80, aproximadamente 80 y aproximadamente 90, aproximadamente 90 y aproximadamente 100, aproximadamente 100 y aproximadamente 150, aproximadamente 150 y aproximadamente 200, aproximadamente 200 y aproximadamente 250, aproximadamente 250 y aproximadamente 300, aproximadamente 300 y aproximadamente 350, aproximadamente 350 y aproximadamente 400, aproximadamente 400 y aproximadamente 450, o aproximadamente 450 y aproximadamente 500 nucleótidos, de manera que los brazos de homología tengan suficiente homología para sufrir una recombinación homóloga con las secuencias diana correspondientes dentro del ácido nucleico diana. Alternativamente, un brazo de homología dado (o cada brazo de homología) y/o la secuencia diana correspondiente pueden comprender regiones de homología correspondientes que tienen una longitud de entre aproximadamente 0,5 kb y aproximadamente 1 kb, aproximadamente 1 kb y aproximadamente 1,5 kb, aproximadamente 1,5 kb y aproximadamente 2 kb, o aproximadamente 2 kb a aproximadamente 2,5 kb. Por ejemplo, los brazos de homología pueden tener cada uno una longitud de aproximadamente 100 nucleótidos. Los brazos de homología pueden ser simétricos (cada uno aproximadamente del mismo tamaño en longitud), o pueden ser asimétricos (uno más largo que el otro).

El casete de borrado puede tener cualquier longitud. Por ejemplo, un casete de eliminación puede ser de aproximadamente 10 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, de aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. En un ejemplo, un casete de eliminación puede tener una longitud de al menos aproximadamente 100 kb o 100 kb. Un casete de eliminación también puede ser de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 100 kb a aproximadamente 125 kb, de aproximadamente 300 kb a aproximadamente 325 kb, de aproximadamente 325 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 375 kb, de aproximadamente 375 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 425 kb, de aproximadamente 425 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 475 kb, o de aproximadamente 475 kb a aproximadamente 500 kb. Alternativamente, un casete de eliminación puede ser de al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más. En un ejemplo, el casete de eliminación tiene una longitud de entre aproximadamente 1 kb y aproximadamente 15 kb o una longitud de aproximadamente 1 kb y aproximadamente 10 kb (por ejemplo, aproximadamente 1,2 kb, aproximadamente 5 kb, aproximadamente 8 kb o aproximadamente 15 kb).

Después de la recombinación homóloga bacteriana, pueden seleccionarse las células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección. Los ejemplos de casetes de selección y métodos de selección se describen con más detalle en otra parte de la presente descripción. En un

ejemplo específico, el casete de selección aporta resistencia a un antibiótico. Por ejemplo, puede aportar resistencia a cualquiera de ampicilina, cloranfenicol, tetraciclina, kanamicina, espectinomicina, estreptomycin, carbenicilina, bleomicina, eritromicina o polimixina B. En algunos métodos, el vector de direccionamiento preexistente también comprende un segundo casete de selección. El segundo casete de selección también puede aportar, por ejemplo, resistencia a un antibiótico. El casete de selección en el casete de eliminación y el segundo casete de selección en el vector de direccionamiento preexistente pueden aportar cada uno resistencia a un antibiótico diferente. Por ejemplo, el casete de selección en el casete de eliminación puede aportar resistencia a un primer antibiótico, y el segundo casete de selección en el vector de direccionamiento preexistente puede aportar resistencia a un segundo antibiótico diferente. En algunos métodos, el segundo casete de selección puede permitir la selección tanto en células bacterianas como en células eucarióticas o de mamíferos.

Después de la selección, el primer sitio diana en el vector de direccionamiento modificado puede escindirse con el primer agente nucleasa, y el segundo sitio diana en el vector de direccionamiento modificado puede escindirse con el segundo agente nucleasa para eliminar el casete de selección y exponer una secuencia final corriente arriba y una secuencia final corriente abajo en el vector de direccionamiento modificado. Por ejemplo, esta etapa puede hacerse *in vitro*. Como ejemplo, el ADN puede aislarse de las células bacterianas después de la recombinación homóloga bacteriana y selección, después de lo cual el primer sitio diana en el vector de direccionamiento modificado puede escindirse con el primer agente nucleasa *in vitro*, y el segundo sitio diana en el vector de direccionamiento modificado puede escindirse con el segundo agente nucleasa *in vitro* para eliminar el casete de selección y exponer la secuencia final corriente arriba y la secuencia final corriente abajo en el vector de direccionamiento modificado.

El primer agente nucleasa y/o el segundo agente nucleasa pueden ser un agente nucleasa de corte raro como se describe en otra parte de la presente descripción. Por ejemplo, en algunos métodos, el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente. Los sitios diana primero y segundo pueden ser diferentes, o el primer sitio diana puede ser idéntico al segundo sitio diana, y el primer agente nucleasa puede ser idéntico al segundo agente nucleasa. El primer agente nucleasa y/o el segundo agente nucleasa pueden crear un extremo romo, una protuberancia 5' o una protuberancia 3'. En un ejemplo, el primer agente nucleasa y/o el segundo agente nucleasa crean una protuberancia 3'.

En un ejemplo específico, el primer agente nucleasa y/o el segundo agente nucleasa es una enzima de restricción o una enzima de restricción de corte raro. Los ejemplos de enzimas de restricción de corte raro se describen en otra parte de la presente descripción, pero pueden incluir, por ejemplo, NotI, XmaIII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AbsI, CciNI, FspAI, MauBI, MreI, MssI, PaeAI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmlI, SrfI, Sse2321, Sse83871, LguI, PciSI, AarI, AjuI, AolI, BarI, PpiI y PstI.

En otro ejemplo específico, el primer agente nucleasa y/o el segundo agente nucleasa pueden ser un agente nucleasa manipulado genéticamente. Por ejemplo, el agente nucleasa puede ser una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (Cas) y un ARN guía (ARNg) (por ejemplo, Cas9 y un ARNg que comprende un ARN CRISPR (ARNcr) y un ARN CRISPR transactivador (ARNtracr)), una nucleasa con dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN) o una meganucleasa manipulada genéticamente.

Después de la escisión/digestión, el vector de direccionamiento escindido puede ensamblarse en una reacción de ensamblaje intermolecular *in vitro* con un casete de modificación que comprende la modificación genética dirigida flanqueada por una secuencia final corriente arriba que se solapa con la secuencia final corriente arriba en el vector de direccionamiento modificado y una secuencia final corriente abajo que se solapa con la secuencia final corriente abajo en el vector de direccionamiento modificado para generar el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices. Por ejemplo, en algunos de tales métodos, ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices.

Puede usarse cualquier método de ensamblaje *in vitro* adecuado. En un ejemplo específico, la etapa de ensamblaje *in vitro* puede comprender incubar el vector de direccionamiento escindido y el casete de modificación con una exonucleasa, una ADN polimerasa y una ADN ligasa. Por ejemplo, el método de ensamblaje *in vitro* puede comprender poner en contacto el vector de direccionamiento escindido y el casete de modificación con una exonucleasa para exponer secuencias complementarias entre las secuencias finales en el vector de direccionamiento modificado y las secuencias finales en el casete de modificación, hibridar las secuencias complementarias expuestas, extender los extremos 3' de las secuencias complementarias hibridadas y ligar las secuencias complementarias hibridadas.

El casete de modificación puede ser un ácido nucleico lineal o un ácido nucleico circular, puede ser un ácido nucleico de cadena única o un ácido nucleico de doble cadena y puede comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). En un ejemplo específico, el casete de modificación es un ADN lineal de doble cadena.

El casete de modificación puede tener cualquier longitud. Por ejemplo, un casete de modificación puede ser de aproximadamente 10 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. En un ejemplo, un casete de modificación puede tener una longitud de al menos aproximadamente 100 kb o 100 kb. Un casete de modificación también puede ser de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 100 kb a aproximadamente 125 kb, de aproximadamente 300 kb a aproximadamente 325 kb, de aproximadamente 325 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 375 kb, de aproximadamente 375 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 425 kb, de aproximadamente 425 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 475 kb, o de aproximadamente 475 kb a aproximadamente 500 kb. Alternativamente, un casete de modificación puede tener al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más. En un ejemplo específico, el casete de modificación tiene una longitud de entre aproximadamente 400 pb y aproximadamente 2 kb. En otro ejemplo, el casete de modificación tiene una longitud de entre aproximadamente 1 kb y aproximadamente 15 kb o de aproximadamente 1 kb y aproximadamente 10 kb (por ejemplo, aproximadamente 1,2 kb, aproximadamente 5 kb, aproximadamente 8 kb o aproximadamente 15 kb). En un ejemplo específico, el casete de modificación tiene una longitud de al menos aproximadamente 200 nucleótidos. En otro ejemplo específico, el casete de modificación tiene un tamaño que no puede sintetizarse o generarse directamente mediante la reacción en cadena de la polimerasa. Por ejemplo, el casete de modificación puede tener una longitud de al menos aproximadamente 5 kb, al menos aproximadamente 10 kb, al menos aproximadamente 15 kb, al menos aproximadamente 20 kb, al menos aproximadamente 25 kb o al menos aproximadamente 30 kb.

La longitud de solapamiento entre la secuencia final corriente arriba en el casete de modificación y la secuencia final corriente arriba en el vector de direccionamiento modificado y/o la longitud de solapamiento entre la secuencia final corriente abajo en el casete de modificación y la secuencia final corriente abajo en el vector de direccionamiento modificado puede tener cualquier longitud adecuada para una reacción de ensamblaje *in vitro*. Como ejemplo, la longitud de solapamiento puede comprender al menos aproximadamente 20 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 40 nucleótidos o al menos aproximadamente 50 nucleótidos. Como otro ejemplo, la longitud de solapamiento puede estar entre aproximadamente 20 nucleótidos y aproximadamente 100 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 90 nucleótidos, entre aproximadamente 0 nucleótidos y aproximadamente 80 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 70 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 60 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 50 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 40 nucleótidos, entre aproximadamente 30 nucleótidos y aproximadamente 60 nucleótidos, o entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos. En un ejemplo específico, la longitud de solapamiento puede estar entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos (por ejemplo, aproximadamente 40 nucleótidos o aproximadamente 50 nucleótidos).

El casete de modificación puede comprender la modificación genética dirigida. Los tipos de modificaciones genéticas dirigidas se describen con más detalle en otra parte de la presente descripción. Algunos ejemplos incluyen mutaciones puntuales, eliminaciones, inserciones, reemplazos o combinaciones de los mismos.

En algunos métodos, para reducir el fondo, el vector producido por el ensamblaje *in vitro* puede tratarse con el primer agente nucleasa y/o el segundo agente nucleasa para reducir el fondo (por ejemplo, mediante la escisión de cualquiera de los vectores que no se ensamblaron exitosamente y por lo tanto aún contenían el sitio diana para el primer agente nucleasa o el segundo agente nucleasa). Tal etapa puede ayudar a verificar que no estén presentes ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa.

#### IV. Recombinación homóloga bacteriana

Cualquier método de recombinación homóloga bacteriana (BHR) adecuado puede usarse en los métodos descritos en la presente descripción. La recombinación homóloga bacteriana implica la expresión transitoria y controlada de genes que median la recombinación homóloga en células bacterianas tales como *Escherichia coli*, lo que permite a las bacterias mediar en la recombinación entre un casete de modificación y un vector de direccionamiento (por ejemplo, un vector de direccionamiento grande) que comparten tramos homólogos cortos. Ver, por ejemplo, el documento US 2004/0018626 y Valenzuela y otros (2003) Nat. Biotechnol. 21(6):652-659.

Los tramos homólogos cortos pueden comprender una región de homología corriente arriba y una región de homología corriente abajo. Las regiones de homología pueden ser de cualquier tamaño adecuado para la recombinación homóloga bacteriana. Por ejemplo, las regiones de homología pueden tener cualquier tamaño adecuado para la recombinación homóloga bacteriana. Por ejemplo, las regiones de homología pueden tener al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 100 nucleótidos. Por ejemplo, las regiones de homología pueden tener entre aproximadamente 35 nucleótidos y 500 nucleótidos, entre aproximadamente 75 nucleótidos y aproximadamente 500 nucleótidos, o entre aproximadamente 50 nucleótidos y aproximadamente 200 nucleótidos (por ejemplo, aproximadamente 100 nucleótidos). Como otro ejemplo, las regiones de homología pueden tener una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 2,5 kb, una longitud de entre aproximadamente 35 nucleótidos y aproximadamente 1,5 kb, o una longitud de entre aproximadamente 35 y aproximadamente 500 nucleótidos. Por ejemplo, una región de homología puede tener una longitud de entre aproximadamente 35 a aproximadamente 40, aproximadamente 40 a aproximadamente 50, aproximadamente 50 a aproximadamente 60, aproximadamente 60 a aproximadamente 70, aproximadamente 70 a aproximadamente 80, aproximadamente 80 a aproximadamente 90, aproximadamente 90 a aproximadamente 100, aproximadamente 100 a aproximadamente 150, aproximadamente 150 a aproximadamente 200, aproximadamente 200 a aproximadamente 250, aproximadamente 250 a aproximadamente 300, aproximadamente 300 a aproximadamente 350, aproximadamente 350 a aproximadamente 400, aproximadamente 400 a aproximadamente 450, o aproximadamente 450 a aproximadamente 500 nucleótidos. Alternativamente, una región de homología dada puede tener una longitud de entre aproximadamente 0,5 kb y aproximadamente 1 kb, aproximadamente 1 kb y aproximadamente 1,5 kb, aproximadamente 1,5 kb y aproximadamente 2 kb, o aproximadamente 2 kb y aproximadamente 2,5 kb. Por ejemplo, la región de homología puede tener una longitud de aproximadamente 100 nucleótidos.

La técnica de modificar un vector de direccionamiento mediante el uso de recombinación homóloga bacteriana puede realizarse en una variedad de sistemas (*ver, por ejemplo*, Yang y otros (1997) *Nat. Biotechnol.* 15:859-65; Muyrers y otros (1999) *Nucleic Acids Res.* 27:1555-1557; Angrand y otros (1999) *Nucleic Acids Res.*, 27:e16; Narayanan y otros (1999) *Gene Ther.*, 6:442-447; y Yu y otros (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:5978-5983). Un ejemplo es la clonación por ET (Zhang y otros (1998) *Nat. Genet.* 20:123-128 y Narayanan y otros (1999) *Gene Ther.*, 6:442-447) y variaciones de esta tecnología (Yu y otros (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:5978-5983). ET se refiere a las proteínas recE y recT que llevan a cabo la reacción de recombinación homóloga. RecE es una exonucleasa que recorta una cadena de ADN lineal de doble cadena de 5' a 3', dejando de este modo un fragmento lineal de doble cadena con una protuberancia de cadena sencilla 3'. Esta protuberancia de cadena única está recubierta por la proteína recT, que tiene actividad de unión al ADN de cadena única (ADNss). La clonación por ET se realiza mediante el uso de *E. coli* que expresan transitoriamente los productos génicos de recE y recT de *E. coli* y la proteína  $\lambda$ gam del bacteriófago lambda ( $\lambda$ ). La proteína  $\lambda$ gam es necesaria para proteger el fragmento de ADN donante de la degradación por el sistema de exonucleasa recBC y es necesaria para la clonación eficiente por ET en huéspedes recBC<sup>+</sup> tales como la cepa DH10b de *E. coli* frecuentemente usada.

#### V. Ensamblaje *in vitro*

Cualquier método de ensamblaje *in vitro* que pueda usarse para ensamblar al menos dos ácidos nucleicos o al menos dos extremos de un solo ácido nucleico en condiciones efectivas para unir las moléculas de ADN para formar una molécula de ADN sustancialmente intacta puede usarse en los métodos descritos en la presente descripción. Algunos ejemplos no limitantes de métodos de ensamblaje *in vitro* incluyen ensamblaje estándar mediante el uso de enzimas de restricción, ensamblaje en fusión, clonación independiente de secuencia y ligasa (SLIC), ensamblaje Gibson y ensamblaje Golden Gate. *Ver, por ejemplo*, Lee y otros (2013) *Mol. Cells* 35:359-370.

Un ejemplo de un método de ensamblaje *in vitro* adecuado es un método isotérmico de reacción única para ensamblar moléculas de ADN de solapamiento mediante la acción concertada de una exonucleasa (por ejemplo, una exonucleasa 5'), una ADN polimerasa y una ADN ligasa. Los ácidos nucleicos que tienen extremos de solapamiento (o un ácido nucleico único lineal con extremos de solapamiento) pueden combinarse con una ligasa, una exonucleasa y una ADN polimerasa. Por ejemplo, dos fragmentos de ADN adyacentes que comparten solapamientos de secuencias terminales pueden unirse en una molécula sellada covalentemente en una reacción isotérmica de una sola etapa. En un ejemplo específico, dos o más moléculas de ADN a ensamblar pueden ponerse en contacto *in vitro* en un solo recipiente con: (a) una exonucleasa 5' a 3' no termoestable aislada que carece de actividad de exonucleasa 3' (por ejemplo, una exonucleasa no procesiva que mastica los extremos de las moléculas de ADN de doble cadena para exponer las protuberancias de una sola cadena que comprenden las regiones de solapamiento); (b) un agente de aglomeración (que, entre otras funciones, puede acelerar la hibridación de ácidos nucleicos, de manera que las protuberancias de cadena únicas se hibridan (aparean) específicamente); (c) una ADN polimerasa termoestable aislada que no desplaza la cadena con actividad exonucleasa 3', o una mezcla de dicha ADN polimerasa con una segunda ADN polimerasa que carece de actividad exonucleasa 3' (para llenar los huecos restantes de una cadena única en las moléculas hibridadas, extendiendo los extremos 3' de las regiones hibridadas); (d) una ligasa termoestable aislada (que sella (liga) las mellas formadas de este modo); (e) una mezcla de dNTP; y (f) un tampón adecuado en condiciones que sean eficaces para unir dos o más moléculas de ADN para formar una

primera molécula de ADNds ensamblada en una reacción de una sola etapa. Para moléculas de cadena única, la exonucleasa puede omitirse, pero no necesariamente. En un ejemplo específico, la exonucleasa T5 elimina los nucleótidos de los extremos 5' de las moléculas de ADN de doble cadena, las protuberancias de ADN de cadena única complementarios se hibridan, y la ADN polimerasa Phusion llena los huecos, y la ADN ligasa Taq sella las mellas. Ver, por ejemplo, los documentos US 2010/0035768, US 2015/0376628, WO 2015/200334, y Gibson y otros (2009) Nat. Methods 6(5):343-345.

Los ácidos nucleicos de cadena única primero y segundo tienen extremos de solapamiento cuando sus respectivos extremos son complementarios entre sí. Los ácidos nucleicos de doble cadena primero y segundo tienen extremos de solapamiento cuando un extremo 5' de una cadena del primer ácido nucleico es complementario al extremo 3' de una cadena del segundo ácido nucleico y viceversa. Por ejemplo, para secuencias de extremos de solapamiento de doble cadena, las cadenas de un ácido nucleico pueden tener al menos aproximadamente 80 %, al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o 100 % de identidad con una cadena correspondiente del otro ácido nucleico. En los métodos descritos en la presente descripción, el extremo 5' de una cadena de una molécula de ADNds que se va a ensamblar comparte secuencias de extremos de solapamiento con el extremo 3' de una cadena de la otra molécula de ADNds. El término secuencias finales de solapamiento incluye ambas cadenas de una molécula de ADNds. Por lo tanto, una cadena de la región de solapamiento puede hibridarse específicamente con su cadena complementaria cuando las regiones complementarias de las secuencias de solapamiento se presentan en protuberancias de cadena única desde los extremos 5' y 3' de los dos polinucleótidos a ensamblar. Puede usarse una exonucleasa para eliminar nucleótidos del extremo 5' o 3' para crear secuencias de extremos protuberantes.

La longitud de la región de solapamiento puede ser lo suficientemente larga como para que la región aparezca solo una vez dentro de cualquiera de los ácidos nucleicos que se están ensamblando. Por lo tanto, se evita que otros polinucleótidos se hibriden con las secuencias finales y el ensamblaje puede ser específico para los ácidos nucleicos diana. Como ejemplo, la longitud de solapamiento puede comprender al menos aproximadamente 20 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 40 nucleótidos o al menos aproximadamente 50 nucleótidos. Como otro ejemplo, la longitud de solapamiento puede estar entre aproximadamente 20 nucleótidos y aproximadamente 100 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 90 nucleótidos, entre aproximadamente 0 nucleótidos y aproximadamente 80 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 70 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 60 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 50 nucleótidos, entre aproximadamente 20 nucleótidos y aproximadamente 40 nucleótidos, entre aproximadamente 30 nucleótidos y aproximadamente 60 nucleótidos, o entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos. En un ejemplo específico, la longitud de solapamiento puede estar entre aproximadamente 40 nucleótidos y aproximadamente 50 nucleótidos (por ejemplo, aproximadamente 40 nucleótidos o aproximadamente 50 nucleótidos).

Las secuencias de solapamiento pueden ponerse en contacto con una exonucleasa para exponer secuencias complementarias (por ejemplo, secuencias complementarias de cadena única) entre las secuencias de solapamiento. La digestión con exonucleasas puede llevarse a cabo en condiciones que sean efectivas para eliminar (masticar) un número suficiente de nucleótidos para permitir la hibridación específica de las regiones de complementariedad de cadena única expuestas. En general, una parte de la región de solapamiento o toda la región de solapamiento se mastica, dejando protuberancias que comprenden una parte de la región de solapamiento o toda la región de solapamiento. En algunos métodos, la digestión con exonucleasas puede llevarse a cabo por una polimerasa en ausencia de dNTP (por ejemplo, ADN polimerasa T5), mientras que en otros métodos la digestión con exonucleasas puede llevarse a cabo por una exonucleasa en presencia de dNTP que carece de actividad polimerasa (por ejemplo, exonucleasa III).

Cualquiera de una variedad de exodesoxirribonucleasas específicas de doble cadena de 5' a 3' puede usarse para masticar los extremos de los ácidos nucleicos en los métodos descritos en la presente descripción. El término exonucleasa 5' se usa algunas veces en la presente descripción para referirse a una exodesoxirribonucleasa 5' a 3'. Una exonucleasa no procesiva se refiere a una exonucleasa que degrada un número limitado de (por ejemplo, sólo unos pocos) nucleótidos durante cada evento de unión al ADN. La digestión con una exonucleasa 5' produce protuberancias de cadena única 3' en las moléculas de ADN. Las exonucleasas 5' usadas en los métodos de ensamblaje *in vitro* pueden carecer de actividad exonucleasa 3', pueden generar extremos de fosfato 5' y pueden iniciar la degradación desde los extremos fosforilados y no fosforilados en 5'. Las exonucleasas usadas en los métodos de ensamblaje *in vitro* descritos en la presente descripción pueden iniciar la digestión desde el extremo 5' de una molécula, ya sea un extremo romo o un pequeño extremo 5' o 3' hundido. Las exonucleasas adecuadas son bien conocidas e incluyen, por ejemplo, exonucleasa de fago T5 (producto del gen D15 del fago T5), exonucleasa del fago lambda, profago RecE de *Rac*, exonucleasa VIII de *E. coli*, exonucleasa del fago T7 (producto del gen 6 del fago T7) o cualquiera de una variedad de exonucleasas 5' que están implicadas en reacciones de recombinación homóloga. Como un ejemplo, la exonucleasa es la exonucleasa T5 o exonucleasa lambda. En un ejemplo específico, la exonucleasa es la exonucleasa T5. En otro ejemplo específico, la exonucleasa no es la exonucleasa de fago T7.

En situaciones en las que la región de solapamiento es larga, puede que solo sea necesario masticar una porción de la región, siempre que las protuberancias de cadena únicas generadas de este modo tengan suficiente longitud y contenido de bases para hibridarse específicamente en las condiciones de la reacción. El término hibridación incluye específicamente situaciones en las que un par particular de protuberancias de cadena únicas se hibridarán preferentemente (o exclusivamente) entre sí, en lugar de otras protuberancias de cadena única (por ejemplo, protuberancias no complementarias) que están presentes en la mezcla de reacción. Por preferencia se entiende que al menos aproximadamente el 95 % de las protuberancias se hibridarán a la protuberancia complementaria. Generalmente, las regiones homólogas de solapamiento (las protuberancias de cadena única o sus complementos) contienen secuencias idénticas. Sin embargo, pueden usarse secuencias parcialmente idénticas, siempre que los protuberancias de cadena únicas puedan hibridarse específicamente en las condiciones de las reacciones.

Después de la hibridación del ADN de cadena única (por ejemplo, protuberancias producidas por la acción de la exonucleasa cuando las moléculas de ADN que se van a unir son ADNds o protuberancias producidas al crear mellas en diferentes sitios diana en cada cadena), los huecos de cadena única dejados por la exonucleasa pueden rellenarse con una polimerasa de ADN adecuada, que no desplaza la cadena, y las mellas formadas de este modo pueden sellarse con una ligasa. Una polimerasa de ADN que no desplaza la cadena, como se usa en la presente descripción, es una polimerasa de ADN que termina la síntesis de ADN cuando encuentra cadenas de ADN que se encuentran en su camino a medida que procede a copiar una molécula de ADNds, o que degrada las cadenas de ADN encontradas a medida que avanza mientras va rellenando al mismo tiempo el hueco creado, generando de este modo una mella en movimiento (traducción de la mella).

Después de hibridar una sola cadena de un primer polinucleótido con la cadena complementaria de un segundo polinucleótido, el extremo 3' del primer polinucleótido puede extenderse en función del molde de la segunda cadena de polinucleótido, y el extremo 3' de la segunda cadena de polinucleótido puede extenderse basándose en la plantilla de la primera cadena de polinucleótidos. Al extender el extremo 3' complementario de cada polinucleótido, pueden ensamblarse los polinucleótidos. Después del ensamblaje, las mellas entre el extremo 3' extendido de una cadena de un fragmento y el extremo 5' adyacente de una cadena del otro fragmento pueden sellarse mediante ligación. Más específicamente, el grupo hidroxilo del extremo 3' extendido del primer polinucleótido puede ligarse al grupo fosfato del extremo 5' del segundo polinucleótido, y el grupo hidroxilo del extremo 3' extendido del segundo polinucleótido puede ligarse al grupo fosfato del extremo 5' del primer polinucleótido.

La reacción de ligación puede realizarse mediante cualquiera de una variedad de ligasas de ADN termoestables adecuadas. Entre las ligasas adecuadas se encuentran, por ejemplo, Taq ligasa, Ampligase Thermostable ADN ligasa, o las ligasas termoestables descritas en la patente de Estados Unidos núm, 6,576,453.

Una cantidad adecuada de un agente de acumulación, tal como PEG, en la mezcla de reacción puede permitir, mejorar o facilitar la acumulación molecular. Tal agente de aglomeración puede permitir que los componentes de la solución entren en contacto más estrecho entre sí. Por ejemplo, las moléculas de ADN que se van a recombinar pueden acercarse más; que puede facilitar la hibridación de las protuberancias de cadena única. Los agentes de apiñamiento adecuados son conocidos e incluyen una variedad de macromoléculas bien conocidas, tales como polímeros como polietilenglicol (PEG), Ficolls tales como Ficoll 70 o dextranos tal como dextrano 70.

Los componentes de la reacción (tales como sales, tampones, una fuente de energía adecuada (tal como ATP o NAD), pH de la mezcla de reacción, etc.) que están presentes en una mezcla de reacción de ensamblaje pueden no ser óptimos para las enzimas individuales (exonucleasa, polimerasa y ligasa), pero pueden servir como un compromiso que es efectivo para todo el conjunto de reacciones.

#### VI. Vectores de direccionamiento y vectores de direccionamiento grandes (LTVEC)

Los vectores de direccionamiento usados en los métodos descritos en la presente descripción pueden ser cualquier vector de direccionamiento adecuado. Los vectores de direccionamiento pueden comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), pueden ser de cadena única o de doble cadena, y pueden tener forma lineal o circular. Los vectores de direccionamiento pueden ser un cromosoma artificial bacteriano (BAC), un BAC modificado o un fragmento de un BAC. Pueden comprender ADN humano, ADN de roedor (por ejemplo, ADN de ratón o ADN de rata), ADN sintético o cualquier combinación de los mismos.

Algunos vectores de direccionamiento usados en los métodos descritos en la presente descripción son vectores de direccionamiento grandes (LTVEC). Los LTVEC incluyen vectores de direccionamiento que comprenden brazos de homología que corresponden y se derivan de secuencias de ácido nucleico más grandes que las usadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en las células. Los LTVEC también incluyen vectores de direccionamiento que comprenden insertos de ácido nucleico que tienen secuencias de ácido nucleico más grandes que las usadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en células. Por ejemplo, los LTVEC hacen posible la modificación de grandes loci que no pueden adaptarse a los vectores de direccionamiento tradicionales basados en plásmidos debido a sus limitaciones de tamaño. Por ejemplo, el locus dirigido puede ser (es decir, los brazos de homología 5' y 3' pueden corresponder a) un locus de la célula que no puede dirigirse como diana mediante el uso de un método convencional o que solo puede dirigirse como

diana incorrectamente o solo con una eficiencia significativamente baja en la ausencia de una mella o ruptura de doble cadena inducida por un agente nucleasa (por ejemplo, una proteína Cas). Los ejemplos de LTVEC incluyen vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o un cromosoma artificial de levadura (YAC). Se describen ejemplos no limitantes de LTVEC y métodos para fabricarlos, por ejemplo, en las patentes de Estados Unidos núms. 6,586,251; 6,596,541; y 7,105,348; y en el documento WO 2002/036789. Los LTVEC pueden estar en forma lineal o en forma circular. Los LTVEC pueden tener cualquier longitud y, típicamente, tienen al menos 10 kb de longitud. El tamaño de un LTVEC puede ser demasiado grande para permitir el tamizaje de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de largo alcance (por ejemplo, de 1 kb a 5 kb).

Los vectores de direccionamiento (por ejemplo, LTVEC) usados en los métodos descritos en la presente descripción pueden tener cualquier longitud. Por ejemplo, un vector de direccionamiento puede ser de aproximadamente 10 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, de aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb o de aproximadamente 350 kb a aproximadamente 400 kb. En un ejemplo, un vector de direccionamiento puede tener al menos aproximadamente 100 kb o 100 kb de longitud. Un vector de direccionamiento también puede ser de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 100 kb a aproximadamente 125 kb, de aproximadamente 300 kb a aproximadamente 325 kb, de aproximadamente 325 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 375 kb, de aproximadamente 375 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 425 kb, de aproximadamente 425 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 475 kb o de aproximadamente 475 kb a aproximadamente 500 kb. Alternativamente, un vector de direccionamiento puede tener al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb o al menos 500 kb o más.

#### VIII. Agentes nucleasa

Cualquier agente nucleasa de corte raro puede usarse en los métodos descritos en la presente descripción. Un agente nucleasa de corte raro es un agente nucleasa con una secuencia diana o una secuencia de reconocimiento que se produce raramente en un genoma. De manera similar, puede usarse cualquier agente nucleasa con una secuencia diana o una secuencia de reconocimiento que no se produzca fuera del(de los) sitio(s) de escisión previsto(s) en los vectores de direccionamiento descritos en la presente descripción. Por ejemplo, puede usarse cualquier agente nucleasa que no tenga una secuencia diana o una secuencia de reconocimiento en los vectores de direccionamiento preexistentes en los métodos descritos en la presente descripción.

Cualquier agente nucleasa como se describe anteriormente que induce una mella o ruptura de doble cadena en una secuencia diana deseada puede usarse en los métodos y composiciones descritos en la presente descripción. Puede emplearse un agente nucleasa natural o nativo siempre que el agente nucleasa induzca una mella o ruptura de doble cadena en una secuencia diana deseada. Alternativamente, puede emplearse un agente nucleasa modificado o manipulado genéticamente. Un "agente nucleasa manipulado genéticamente" incluye una nucleasa que se manipula (modifica o deriva) a partir de su forma nativa para reconocer e inducir específicamente una mella o ruptura de doble cadena en la secuencia diana deseada. Por lo tanto, un agente nucleasa manipulado genéticamente puede derivarse de un agente nucleasa nativo de origen natural o puede crearse o sintetizarse artificialmente. La nucleasa manipulada genéticamente puede inducir una mella o ruptura de doble cadena en una secuencia diana, por ejemplo, en donde la secuencia diana no es una secuencia que habría sido reconocida por un agente nucleasa nativo (no manipulado genéticamente o no modificado). La modificación del agente nucleasa puede ser tan pequeña como un aminoácido en un agente de escisión de proteínas o un nucleótido en un agente de escisión de ácido nucleico. La producción de una mella o ruptura de doble cadena en una secuencia diana u otro ADN puede denominarse en la presente descripción "cortar" o "escindir" la secuencia diana u otro ADN.

También se proporcionan variantes activas y fragmentos de las secuencias diana ejemplificadas. Tales variantes activas pueden comprender al menos 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de identidad de secuencia con la secuencia diana dada, en donde las variantes activas conservan la actividad biológica y, por lo tanto, pueden ser reconocidas y escindidas por un agente nucleasa de una manera específica de secuencia. Los ensayos para medir la ruptura de doble cadena de una secuencia diana por un agente nucleasa son bien conocidos. *Ver, por ejemplo*, Frenthewey y otros (2010) *Methods in Enzymology* 476:295-307.

También se proporcionan variantes activas y fragmentos de agentes nucleasa (es decir, un agente nucleasa manipulado genéticamente). Tales variantes activas pueden comprender al menos un 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de identidad de secuencia con el agente nucleasa nativo, en donde las variantes activas conservan la capacidad de cortar en una secuencia diana deseada y, por lo tanto, conservan la actividad inductora de mella o ruptura de doble cadena. Por ejemplo, cualquiera de los agentes nucleasa descritos en la presente descripción puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir una mella o ruptura de doble cadena en una secuencia diana que no fue reconocida por el agente nucleasa nativo. Por lo tanto, algunas nucleasas manipuladas genéticamente tienen una especificidad para inducir una mella o ruptura de doble cadena en una secuencia diana que es diferente de la secuencia diana del agente nucleasa nativo correspondiente. Los ensayos para determinar la actividad inductora de mella o ruptura de doble cadena son conocidos y generalmente miden la actividad global y la especificidad de la endonucleasa sobre sustratos de ADN que contienen la secuencia diana.

Una secuencia diana de nucleasa incluye una secuencia de ADN en la que un agente nucleasa induce una mella o ruptura de doble cadena. La longitud de la secuencia diana puede variar e incluye, por ejemplo, secuencias diana que tienen aproximadamente 30-36 pb para un par de nucleasas con dedos de zinc (ZFN) (es decir, aproximadamente 15-18 pb para cada ZFN), aproximadamente 36 pb para una nucleasa efectora de tipo activador de la transcripción (TALEN), o aproximadamente 20 pb para un ARN guía de CRISPR/Cas9.

#### A. Enzimas de restricción

Los agentes de nucleasa adecuados para usar en los métodos descritos en la presente descripción pueden comprender endonucleasas de restricción, que incluyen endonucleasas de Tipo I, Tipo II, Tipo III y Tipo IV. Las endonucleasas de restricción tipo I y tipo III reconocen sitios de reconocimiento específicos, pero típicamente se escinden en una posición variable desde el sitio de unión de la nucleasa, que puede estar a cientos de pares de bases del sitio de escisión (sitio de reconocimiento). En los sistemas de tipo II, la actividad de restricción es independiente de cualquier actividad de metilasa, y la escisión típicamente se produce en sitios específicos dentro o cerca del sitio de unión. La mayoría de las enzimas de tipo II cortan secuencias palindrómicas, sin embargo, las enzimas de tipo IIa reconocen sitios de reconocimiento no palindrómicos y se escinden fuera del sitio de reconocimiento, las enzimas de tipo IIb cortan secuencias dos veces con ambos sitios fuera del sitio de reconocimiento y las enzimas de tipo II reconocen un sitio de reconocimiento asimétrico y escinden por un lado ya una distancia definida de aproximadamente 1-20 nucleótidos del sitio de reconocimiento. Las enzimas de restricción de tipo IV se dirigen al ADN metilado. Las enzimas de restricción se describen y clasifican con más detalle, por ejemplo, en la base de datos REBASE (página web en [rebase.neb.com](http://rebase.neb.com); Roberts y otros (2003) *Nucleic Acids Res.* 31:418-20); Roberts y otros (2003) *Nucleic Acids Res.* 31:1805-12; y Belfort y otros (2002) en *Mobile DNA II*, páginas 761-783, Eds. Craigie y otros (ASM Press, Washington, DC).

En algunos métodos, se usa una enzima de restricción de corte raro. Una enzima de restricción de corte raro se refiere a una enzima con un sitio diana o un sitio de reconocimiento que se produce rara vez en un genoma. El tamaño de los fragmentos de restricción generados al cortar un genoma aleatorio hipotético con una enzima de restricción puede aproximarse a  $4^N$ , donde N es el número de nucleótidos en el sitio de reconocimiento de la enzima. Por ejemplo, una enzima con un sitio de reconocimiento que consiste en 7 nucleótidos cortaría un genoma una vez cada  $4^7$  pb, produciendo fragmentos de aproximadamente 16 384 pb. Generalmente, las enzimas de corte raro tienen sitios de reconocimiento que comprenden 6 o más nucleótidos. Por ejemplo, una enzima de corte raro puede tener un sitio de reconocimiento que comprende o consiste en 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 nucleótidos. Los ejemplos de enzimas de restricción de corte raro incluyen NotI (GCGGCCGC), XmaII (CGGCCG), SstII (CCGCGG), Sall (GTCGAC), NruI (TCGCGA), NheI (GCTAGC), Nb.BbvCI (CCTCAGC), BbvCI (CCTCAGC), AscI (GGCGGCC), AsiSI (GCGATCGC), FseI (GGCCGGCC), PaeI (TTAATTAA), PmeI (GTTTAAAC), SbfI (CCTGCAGG), SgrAI (CRCCGGYG), SwaI (ATTTAAAT), BspQI (GCTCTTC), SapI (GCTCTTC), SfiI (GGCCNNNNNGGCC), CspCI (CAANNNNNGTGG), AbsI (CCTCGAGG), CciNI (GCGGCCGC), FspAI (RTGCGCAY), MauBI (CGCGCGCG), MreI (CGCCGGCG), MssI (GTTTAAAC), PaeI (GGCGGCC), RgaI (GCGATCGC), RgiI (GGCCGGCC), SdaI (CCTGCAGG), SfaI (GCGATCGC), SgfI (GCGATCGC), SgrDI (CGTCGACG), SgsI (GGCGGCC), SmiI (ATTTAAAT), SrfI (GCCGGGC), Sse2321 (CGCCGGCG), Sse83871 (CCTGCAGG), LglI (GCTCTTC), PciSI (GCTCTTC), AarI (CACCTGC), AjuI (GAANNNNNNNTTGG), AfoI (GAACNNNNNTCC), BarI (GAAGNNNNNTAC), PpiI (GAACNNNNNCTC), PstI (GAACNNNNNTAC) y otras.

#### B. Sistemas CRISPR/Cas

Los sistemas de repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR)/asociados a CRISPR (Cas) también pueden usarse como agentes de nucleasa de corte raro en los métodos descritos en la presente descripción. Los sistemas CRISPR/Cas incluyen transcripciones y otros elementos involucrados en la expresión o dirección de la actividad de los genes Cas. Un sistema CRISPR/Cas puede ser, por ejemplo, un sistema de tipo I, tipo II, tipo III o tipo V (por ejemplo, subtipo VA o subtipo VB). Los sistemas CRISPR/Cas usados en las composiciones y los métodos descritos en la presente descripción pueden no producirse de forma natural. Un sistema de "origen no natural" incluye cualquier cosa que indique la participación de la mano del hombre, tal como uno o más componentes del sistema alterados o mutados de su estado de origen natural, estando al menos

5 sustancialmente libre de al menos otro componente con los que están naturalmente asociados en la naturaleza, o están asociados con al menos otro componente con el que no están naturalmente asociados. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR de origen no natural que comprenden un ARNg y una proteína Cas que no se producen juntos de forma natural, emplean una proteína Cas que no se produce de forma natural o emplean un ARNg que no se produce de forma natural.

10 Proteínas Cas y polinucleótidos que codifican proteínas Cas. Las proteínas Cas generalmente comprenden al menos un dominio de unión o reconocimiento de ARN que puede interactuar con los ARN guía (ARNg). Las proteínas Cas también pueden comprender dominios de nucleasa (por ejemplo, dominios de DNasa o dominios de RNasa), dominios de unión a ADN, dominios de helicasa, dominios de interacción proteína-proteína, dominios de dimerización y otros dominios. Algunos de tales dominios (por ejemplo, dominios de ADNasa) pueden ser de una proteína Cas nativa. Pueden añadirse otros dominios de este tipo para hacer una proteína Cas modificada. Un dominio de nucleasa posee actividad catalítica para la escisión de ácidos nucleicos, que incluye la ruptura de los enlaces covalentes de una molécula de ácido nucleico. La escisión puede producir extremos romos o extremos escalonados, y puede ser de cadena única o de doble cadena. Por ejemplo, una proteína Cas9 de tipo silvestre típicamente creará un producto de escisión romo. Alternativamente, una proteína Cpf1 de tipo silvestre (por ejemplo, FnCpf1) puede dar como resultado un producto de escisión con una protuberancia 5' de 5 nucleótidos, y la escisión se produce después del par de bases 18 de la secuencia PAM en la cadena no dirigida y después de la base 23 en la cadena diana. Una proteína Cas puede tener una actividad de escisión completa para crear una ruptura de doble cadena en un locus genómico diana (por ejemplo, una ruptura de doble cadena con extremos romos), o puede ser una nickasa que crea una ruptura de cadena única en un locus genómico diana.

25 Los ejemplos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 o Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), 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 y Cu1966, y homólogos o versiones modificadas de los mismos.

30 Una proteína Cas ilustrativa es una proteína Cas9 o una proteína derivada de una proteína Cas9. Las proteínas Cas9 son de un sistema CRISPR/Cas de tipo II y típicamente comparten cuatro motivos clave con una arquitectura conservada. Los motivos 1, 2 y 4 son motivos similares a RuvC y el motivo 3 es un motivo HNH. Las proteínas Cas9 ilustrativas son de *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Nocardiaopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor becscii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleia chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, *Acaryochloris marina*, *Neisseria meningitidis* o *Campylobacter jejuni*. Se describen ejemplos adicionales de los miembros de la familia Cas9 en el documento WO 2014/131833. Cas9 de *S. pyogenes* (SpCas9) (número de acceso de SwissProt asignado Q99ZW2) es una proteína Cas9 ilustrativa. Cas9 de *S. aureus* (SaCas9) (número de acceso de UniProt asignado J7RUA5) es otra proteína Cas9 ilustrativa. Cas9 de *Campylobacter jejuni* (CjCas9) (número de acceso de UniProt asignado Q0P897) es otra proteína Cas9 ilustrativa. Ver, por ejemplo, Kim y otros (2017) Nat. Commun. 8:14500. SaCas9 es más pequeño que SpCas9 y CjCas9 es más pequeño que SaCas9 y SpCas9. Cas9 de *Neisseria meningitidis* (Nme2Cas9) es otra proteína Cas9 ilustrativa. Ver, por ejemplo, Edraki y otros (2019) Mot. Cell 73(4):714-726. Las proteínas Cas9 de *Streptococcus thermophilus* (por ejemplo, LMD-9 Cas9 de *Streptococcus thermophilus* codificada por el locus CRISPR1 (St1Cas9) o Cas9 de *Streptococcus thermophilus* del locus CRISPR3 (St3Cas9)) son otras proteínas Cas9 ilustrativas. Cas9 de *Francisella novicida* (FnCas9) o la variante Cas9 RHA de *Francisella novicida* que reconoce una PAM alternativa (sustituciones E1369R/E1449H/R1556A) son otras proteínas Cas9 ilustrativas. Estas y otras proteínas Cas9 ilustrativas se revisan, por ejemplo, en Cebrian-Serrano y Davies (2017) Mamm. Genome 28(7):247-261. Una secuencia de proteína Cas9 ilustrativa puede comprender, consistir esencialmente en, o consistir en la SEQ ID NO: 1. Un ADN ilustrativo que codifica la proteína Cas9 puede comprender, consistir esencialmente en, o consistir en la SEQ ID NO: 2.

60 Otro ejemplo de una proteína Cas es una proteína Cpf1 (CRISPR de *Prevotella* y *Francisella* 1). Cpf1 es una proteína grande (aproximadamente 1300 aminoácidos) que contiene un dominio de nucleasa similar a RuvC homólogo al dominio correspondiente de Cas9 junto con una contraparte del característico grupo rico en arginina de Cas9. Sin embargo, Cpf1 carece del dominio de nucleasa HNH que está presente en las proteínas Cas9, y el dominio similar a RuvC es contiguo en la secuencia de Cpf1, en contraste con Cas9, donde contiene insertos largos que incluyen el dominio HNH. Ver, por ejemplo, Zetsche y otros (2015) Cell 163(3):759-771. Las proteínas Cpf1 ilustrativas son de *Francisella tularensis* 1, *Francisella tularensis subsp. novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium*

MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium GW2011\_GWA2\_33\_10*, *Parcubacteria bacterium GW2011\_GWC2\_44\_17*, *Smithella sp. SCADC*, *Acidaminococcus sp. BV3L6*, *Lachnospiraceae bacterium MA2020*, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi 237*, *Leptospira inadai*, *Lachnospiraceae bacterium ND2006*, *Porphyromonas crevioricanis 3*, *Prevotella disiens* y *Porphyromonas macacae*. La Cpf1 de *Francisella novicida* U112 (FnCpf1; número de acceso de UniProt asignado A0Q7Q2) es una proteína Cpf1 ilustrativa.

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, las que se producen en la naturaleza), proteínas Cas modificadas (es decir, variantes de proteína Cas) o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las proteínas Cas también pueden ser variantes o fragmentos activos con respecto a la actividad catalítica de las proteínas Cas de tipo silvestre o modificadas. Las variantes o fragmentos activos con respecto a la actividad catalítica pueden comprender al menos un 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de identidad de secuencia con la proteína Cas de tipo silvestre o modificada o una porción de la misma, en donde las variantes activas conservan la capacidad de cortar en un sitio de escisión deseado y por lo tanto conservan la actividad inductora de mella o inductora de ruptura de doble cadena. Se conocen ensayos para la actividad inductora de mella o inductora de ruptura de doble cadena y, en general, miden la actividad global y la especificidad de la proteína Cas sobre sustratos de ADN que contienen el sitio de escisión.

Las proteínas Cas pueden modificarse para aumentar o disminuir uno o más de la afinidad de unión de ácidos nucleicos, la especificidad de unión de ácidos nucleicos y la actividad enzimática. Las proteínas Cas también pueden modificarse para cambiar cualquier otra actividad o propiedad de la proteína, tal como la estabilidad. Por ejemplo, uno o más dominios de nucleasa de la proteína Cas pueden modificarse, eliminarse o inactivarse, o una proteína Cas puede truncarse para eliminar dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad o una propiedad de la proteína Cas.

Un ejemplo de una proteína Cas modificada es la proteína SpCas9-HF1 modificada, que es una variante de alta fidelidad de Cas9 de *Streptococcus pyogenes* que alberga alteraciones (N497A/R661A/Q695A/Q926A) diseñada para reducir los contactos de ADN no específicos. Ver, por ejemplo, Kleinstiver y otros (2016) Nature 529(7587):490-495. Otro ejemplo de una proteína Cas modificada es la variante eSpCas9 modificada (K848A/K1003A/R1060A) diseñada para reducir los efectos secundarios. Ver, por ejemplo, Slaymaker y otros (2016) Science 351(6268):84-88. Otras variantes de SpCas9 incluyen K855A y K810A/K1003A/R1060A. Estas y otras proteínas Cas modificadas se revisan, por ejemplo, en Cebrian-Serrano y Davies (2017) Mamm. Genome 28(7):247-261. Otro ejemplo de una proteína Cas9 modificada es xCas9, que es una variante de SpCas9 que puede reconocer una gama ampliada de secuencias PAM. Ver, por ejemplo, Hu y otros (2018) Nature 556:57-63.

Las proteínas Cas pueden comprender al menos un dominio de nucleasa, tal como un dominio de DNasa. Por ejemplo, una proteína Cpf1 de tipo silvestre generalmente comprende un dominio similar a RuvC que escinde ambas cadenas de ADN diana, quizás en una configuración dimerica. Las proteínas Cas también pueden comprender al menos dos dominios de nucleasa, tal como los dominios de DNasa. Por ejemplo, una proteína Cas9 de tipo silvestre generalmente comprende un dominio de nucleasa de tipo RuvC y un dominio de nucleasa de tipo HNH. Cada uno de los dominios RuvC y HNH puede cortar una cadena diferente de ADN de doble cadena para hacer una ruptura de doble cadena en el ADN. Ver, por ejemplo, Jinek y otros (2012) Science 337:816-821.

Uno o más de los dominios de nucleasa pueden eliminarse o mutarse para que ya no sean funcionales o tengan una actividad nucleasa reducida. Por ejemplo, si uno de los dominios de nucleasa se elimina o muta en una proteína Cas9, la proteína Cas9 resultante puede denominarse nickasa y puede generar una ruptura de cadena única dentro de un ADN diana de doble cadena, pero no una ruptura de doble cadena (es decir, puede escindir la cadena complementaria o la no complementaria, pero no ambas). Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato a alanina en la posición 10 de Cas9) en el dominio RuvC de Cas9 de *S. pyogenes*. Asimismo, H939A (histidina a alanina en la posición de aminoácido 839), H840A (histidina a alanina en la posición de aminoácido 840) o N863A (asparagina a alanina en la posición de aminoácido N863) en el dominio HNH de Cas9 de *S. pyogenes* puede convertir el Cas9 en una nickasa. Otros ejemplos de mutaciones que convierten a Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Ver, por ejemplo, Sapranauskas y otros (2011) Nucleic Acids Res. 39(21):9275-9282 y el documento WO 2013/141680. Tales mutaciones pueden generarse mediante el uso de métodos tales como la mutagénesis dirigida al sitio, la mutagénesis mediada por PCR o la síntesis total de genes. Pueden encontrarse ejemplos de otras mutaciones que crean nickasas, por ejemplo, en los documentos WO 2013/176772 y WO 2013/142578,

Los ejemplos de mutaciones inactivantes en los dominios catalíticos de xCas9 son los mismos que los descritos anteriormente para SpCas9. También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de las proteínas Cas9 de *Staphylococcus aureus*. Por ejemplo, la enzima Cas9 de *Staphylococcus aureus* (SaCas9) puede comprender una sustitución en la posición N580 (por ejemplo, sustitución N580A) para crear una nickasa. Alternativamente, la enzima SaCas9 puede comprender una sustitución en la posición D10 (por ejemplo, sustitución D10A) para generar una nickasa. Ver, por ejemplo, el documento WO 2016/106236. También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de Nme2Cas9 (por ejemplo, combinación de D16A y H588A). También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de St1Cas9 (por ejemplo,

combinación de D9A, D598A, H599A y N622A). También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de St3Cas9 (por ejemplo, combinación de D10A y N870A). También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de CjCas9 (por ejemplo, combinación de D8A y H559A). También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de FnCas9 y RHA FnCas9 (por ejemplo, N995A).

También se conocen ejemplos de mutaciones inactivantes en los dominios catalíticos de las proteínas Cpf1. Con referencia a las proteínas Cpf1 de *Francisella novicida* U112 (FnCpf1), *Acidaminococcus* sp. BV3L6 (AsCpf1), *Lachnospiraceae bacterium* ND2006 (LbCpf1) y *Moraxella bovoculi* 237 (MbCpf1 Cpf1), tales mutaciones pueden incluir mutaciones en las posiciones 908, 993 o 1263 de AsCpf1 o las posiciones correspondientes en los ortólogos de Cpf1, o las posiciones 832, 925, 947 o 1180 de LbCpf1 o las posiciones correspondientes en los ortólogos de Cpf1. Tales mutaciones pueden incluir, por ejemplo, una o más mutaciones D908A, E993A y D1263A de AsCpf1 o mutaciones correspondientes en ortólogos de Cpf1, o D832A, E925A, D947A y D1180A de LbCpf1 o mutaciones correspondientes en ortólogos de Cpf1. Ver, por ejemplo, el documento US 2016/0208243.

Las proteínas Cas también pueden unirse operativamente a polipéptidos heterólogos como proteínas de fusión. Por ejemplo, una proteína Cas puede fusionarse con un dominio de escisión. Ver el documento WO 2014/089290. Las proteínas Cas también pueden fusionarse con un polipéptido heterólogo proporcionando una mayor o menor estabilidad. El dominio fusionado o polipéptido heterólogo puede localizarse en el extremo N-terminal, el extremo C-terminal o internamente dentro de la proteína Cas.

Como un ejemplo, una proteína Cas puede fusionarse con uno o más polipéptidos heterólogos que proporcionan localización subcelular. Tales polipéptidos heterólogos pueden incluir, por ejemplo, una o más señales de localización nuclear (NLS) tal como la SV40 NLS monopartita y/o una NLS bipartita unida a alfa-importina para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a las mitocondrias, una señal de retención de ER, y similares. Ver, por ejemplo, Lange y otros (2007) J. Biol. Chem. 282(8):5101-5105. Tales señales de localización subcelular pueden localizarse en el extremo N-terminal, el extremo C-terminal o en cualquier lugar dentro de la proteína Cas. Una NLS puede comprender un tramo de aminoácidos básicos y puede ser una secuencia monopartita o una secuencia bipartita. Opcionalmente, una proteína Cas puede comprender dos o más NLS, incluida una NLS (por ejemplo, una NLS unida a alfa-importina o una NLS monopartita) en el extremo N-terminal y una NLS (por ejemplo, una V40 NLS o una NLS bipartita) en el extremo C-terminal. Una proteína Cas también puede comprender dos o más NLS en el extremo N-terminal y/o dos o más NLS en el extremo C-terminal.

Las proteínas Cas también pueden unirse operativamente a un dominio de penetración celular o a un dominio de transducción de proteínas. Por ejemplo, el dominio de penetración celular puede derivar de la proteína TAT del VIH-1, el motivo de penetración celular TLM del virus de la hepatitis B humana, MPG, Pep-1, VP22, un péptido de penetración celular del virus Herpes simplex o un secuencia peptídica de poliarginina. Ver, por ejemplo, los documentos WO 2014/089290 y WO 2013/176772. El dominio de penetración celular puede localizarse en el extremo N-terminal, el extremo C-terminal o en cualquier lugar dentro de la proteína Cas.

Las proteínas Cas proporcionadas como ARNm pueden modificarse para mejorar la estabilidad y/o las propiedades de inmunogenicidad. Las modificaciones pueden realizarse en uno o más nucleósidos dentro del ARNm. Los ejemplos de modificaciones químicas de las nucleobases de ARNm incluyen pseudouridina, 1-metil-pseudouridina y 5-metil-citidina. Por ejemplo, puede usarse ARNm de Cas protegido y poliadenilado que contiene N1-metil pseudouridina. Asimismo, los ARNm de Cas pueden modificarse mediante el agotamiento de la uridina mediante el uso de codones sinónimos.

ARN guía. Un "ARN guía" o "ARNg" es una molécula de ARN que se une a una proteína Cas (por ejemplo, la proteína Cas9) y dirige la proteína Cas a una localización específica dentro de un ADN diana. Los ARN guía pueden comprender dos segmentos: un "segmento de direccionamiento al ADN" y un "segmento de unión a proteínas". "Segmento" incluye una sección o región de una molécula, tal como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg, tales como los de Cas9, pueden comprender dos moléculas de ARN separadas: un "ARN activador" (por ejemplo, ARNtracr) y un "ARN direccionador" (por ejemplo, ARN CRISPR o ARNcr). Otros ARNg son una sola molécula de ARN (polinucleótido de ARN único), que también pueden denominarse "ARNg de molécula única", "ARN guía único" o "ARNsg". Ver, por ejemplo, los documentos WO 2013/176772, WO 2014/065596, WO 2014/089290, WO 2014/093622, WO 2014/099750, WO 2013/142578 y WO 2014/131833. Para Cas9, por ejemplo, un ARN guía único puede comprender un ARNcr fusionado con un ARNtracr (por ejemplo, a través de un enlace). Para Cpf1, por ejemplo, solo se necesita un ARNcr para lograr la unión y/o la escisión de una secuencia diana. Los términos "ARN guía" y "ARNg" incluyen tanto ARNg de molécula doble (es decir, modulares) como ARNg de molécula única.

Un ARNg de dos moléculas ilustrativas comprende una molécula similar a ARNcr ("ARN CRISPR" o "ARN direccionador" o "ARNcr" o "repetición de ARNcr") y una correspondiente molécula similar a ARNtracr ("ARN CRISPR transactivador" o "ARN activador" o "ARNtracr"). Un ARNcr comprende tanto el segmento de direccionamiento al ADN (de cadena única) del ARNg como un tramo de nucleótidos (es decir, la cola del ARNcr) que forma la mitad del dúplex de ARNds del segmento de unión a proteínas del ARNg. Un ejemplo de una cola de

ARNcr, localizada corriente abajo (3') del segmento de direccionamiento al ADN, comprende, consiste esencialmente en, o consiste en GUUUUAGAGCUAUGCU (SEQ ID NO: 3). Cualquiera de los segmentos de direccionamiento al ADN descritos en la presente descripción puede unirse al extremo 5' de SEQ ID NO: 3 para formar un ARNcr.

Un ARNtracr (ARN activador) correspondiente comprende un tramo de nucleótidos que forma la otra mitad del dúplex de ARNds del segmento de unión a proteínas del ARNg. Un tramo de nucleótidos de un ARNcr son complementarios y se hibridan con un tramo de nucleótidos de un ARNtracr para formar el dúplex de ARNds del dominio de unión a proteínas del ARNg. Como tal, puede decirse que cada ARNcr tiene un ARNtracr correspondiente. Un ejemplo de una secuencia de ARNtracr comprende, consiste esencialmente en o consiste en AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACC GAGUCGGUGCUUU (SEQ ID NO: 4). Otros ejemplos de secuencias de ARNtracr comprenden, consisten esencialmente en, o consisten en

AAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG

CACCGAGUCGGUGCUUUU (SEQ ID NO: 12)

GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA  
ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 13).

En los sistemas en los que se necesitan tanto un ARNcr como un ARNtracr, el ARNcr y el ARNtracr correspondiente se hibridan para formar un ARNg. En los sistemas en los que solo se necesita un ARNcr, el ARNcr puede ser el ARNg. El ARNcr proporciona adicionalmente el segmento de direccionamiento al ADN de cadena única que se hibrida con la cadena complementaria de un ADN diana. Si se usa para la modificación dentro de una célula, la secuencia exacta de una molécula de ARNcr o ARNtracr determinada puede diseñarse para que sea específica de la especie en la que se usarán las moléculas de ARN. Ver, por ejemplo, Mali et al. (2013) Science 339(6121):823-826; Jinek y otros (2012) Science 337(6096):816-821; Hwang y otros (2013) Nat. Biotechnol. 31(3):227-229; Jiang y otros (2013) Nat. Biotechnol 31(3):233-239; y Cong y otros (2013) Science 339(6121):819-823.

El segmento de direccionamiento al ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en la cadena complementaria del ADN diana, como se describe con más detalle a continuación. El segmento de direccionamiento al ADN de un ARNg interactúa con el ADN diana de una manera específica de secuencia a través de la hibridación (es decir, apareamiento de bases). Como tal, la secuencia de nucleótidos del segmento de ADN diana puede variar y determina la localización dentro del ADN diana con el que interactuarán el ARNg y el ADN diana. El segmento de direccionamiento al ADN de un ARNg sujeto puede modificarse para hibridar con cualquier secuencia deseada dentro de un ADN diana. Los ARNcr de origen natural difieren según el sistema CRISPR/Cas y el organismo, pero frecuentemente contienen un segmento de direccionamiento de entre 21 y 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (ver, por ejemplo, el documento WO 2014/131833-). En el caso de *S. pyogenes*, las DR tienen una longitud de 36 nucleótidos y el segmento de direccionamiento tiene una longitud de 30 nucleótidos. La DR localizada en 3' es complementaria y se hibrida con el ARNtracr correspondiente, que a su vez se une a la proteína Cas.

El segmento de direccionamiento al ADN puede tener, por ejemplo, una longitud de al menos aproximadamente 12, 15, 17, 18, 19, 20, 25, 30, 35 o 40 nucleótidos. Tales segmentos de direccionamiento al ADN pueden tener, por ejemplo, una longitud de aproximadamente 12 a aproximadamente 100, de aproximadamente 12 a aproximadamente 80, de aproximadamente 12 a aproximadamente 50, de aproximadamente 12 a aproximadamente 40, de aproximadamente 12 a aproximadamente 30, de aproximadamente 12 a aproximadamente 25 o de aproximadamente 12 a aproximadamente 20 nucleótidos. Por ejemplo, el segmento de direccionamiento al ADN puede ser de aproximadamente 15 a aproximadamente 25 nucleótidos (por ejemplo, de aproximadamente 17 a aproximadamente 20 nucleótidos, o de aproximadamente 17, 18, 19 o 20 nucleótidos). Ver, por ejemplo, el documento US 2016/0024523. Para Cas9 de *S. pyogenes*, un segmento de direccionamiento al ADN típico tiene una longitud de entre 16 y 20 nucleótidos o una longitud de entre 17 y 20 nucleótidos. Para Cas9 de *S. aureus*, un segmento de direccionamiento al ADN típico tiene una longitud de entre 21 y 23 nucleótidos. Para Cpf1, un segmento de direccionamiento al ADN típico tiene una longitud de al menos 16 nucleótidos de longitud o una longitud de al menos 18 nucleótidos.

Los ARNtracr pueden estar en cualquier forma (por ejemplo, ARNtracr completos o ARNtracr parciales activos) y ser de diferentes longitudes. Pueden incluir transcripciones primarias o formularios procesados. Por ejemplo, los ARNtracr (como parte de un ARN guía único o como una molécula separada como parte de un ARNg de dos moléculas) pueden comprender, consistir esencialmente en, o consistir en la totalidad o una porción de una secuencia de ARNtracr de tipo silvestre (por ejemplo, aproximadamente o más de aproximadamente 20, 26, 32, 45,

5 48, 54, 63, 67, 85 o más nucleótidos de una secuencia de ARNtracr de tipo silvestre). Los ejemplos de secuencias de ARNtracr de tipo silvestre de *S. pyogenes* incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos y 65 nucleótidos. Ver, por ejemplo, Deltcheva y otros (2011) Nature 471(7340):602-607; WO 2014/093661. Los ejemplos de ARNtracr dentro de los RNA guía únicos (ARNgs) incluyen los segmentos de ARNtracr que se encuentran en las versiones +48, +54, +67 y +85 de los ARNgs, donde "+n" indica que hasta el nucleótido +n del ARNtracr de tipo silvestre está incluido en el ARNgs. Ver el documento US 8,697,359.

10 El porcentaje de complementariedad entre el segmento de direccionamiento al ADN del ARN guía y la cadena complementaria del ADN diana puede ser de al menos 60 % (por ejemplo, al menos 65 %, al menos 70 %, al menos 75 %, al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 %). El porcentaje de complementariedad entre el segmento de direccionamiento al ADN y la cadena complementaria del ADN diana puede ser de al menos 60 % sobre aproximadamente 20 nucleótidos contiguos. Como ejemplo, el porcentaje de complementariedad entre el segmento de direccionamiento al ADN y la cadena complementaria del ADN diana puede ser del 100 % sobre los 14 nucleótidos contiguos en el extremo 5' de la cadena complementaria del ADN diana y tan bajo como del 0 % sobre el resto. En tal caso, puede considerarse que el segmento de direccionamiento al ADN tiene una longitud de 14 nucleótidos. Como otro ejemplo, el porcentaje de complementariedad entre el segmento de direccionamiento al ADN y la cadena complementaria del ADN diana puede ser del 100 % sobre los siete nucleótidos contiguos en el extremo 5' de la cadena complementaria del ADN diana y tan bajo como 0 % sobre el resto. En tal caso, puede considerarse que el segmento de direccionamiento al ADN tiene una longitud de 7 nucleótidos. En algunos ARN guía, al menos 17 nucleótidos dentro del segmento de direccionamiento al ADN son complementarios a la cadena complementaria del ADN diana. Por ejemplo, el segmento de direccionamiento al ADN puede tener una longitud de 20 nucleótidos y puede comprender 1, 2 o 3 errores de apareamiento con la cadena complementaria del ADN diana. En un ejemplo, los errores de apareamiento no son adyacentes a la región de la cadena complementaria correspondiente a la secuencia del motivo adyacente del protoespaciador (PAM) (es decir, el complemento inverso de la secuencia de PAM) (por ejemplo, los errores de apareamiento están en el extremo 5' del segmento de direccionamiento al ADN del ARN guía, o los errores de apareamiento son al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 o 19 pares de bases de distancia de la región de la cadena complementaria correspondiente a la secuencia de PAM).

30 El segmento de unión a proteínas de un ARNg puede comprender dos tramos de nucleótidos que son complementarios entre sí. Los nucleótidos complementarios del segmento de unión a proteínas se hibridan para formar un dúplex de ARN de doble cadena (ARNds). El segmento de unión a proteínas de un ARNg sujeto interactúa con una proteína Cas, y el ARNg dirige la proteína Cas unida a una secuencia de nucleótidos específica dentro del ADN diana a través del segmento de direccionamiento al ADN.

35 Los ARN guía únicos pueden comprender un segmento de direccionamiento al ADN unido a una secuencia de andamio (es decir, la secuencia de unión a proteína o de unión a Cas del ARN guía). Por ejemplo, tales ARN guía pueden tener un segmento de direccionamiento al ADN 5' y una secuencia de andamio 3'. Las secuencias de andamio ilustrativas comprenden, consisten esencialmente en o consisten en:

40 GUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCAACUUGA  
AAAAGUGGCACCGAGUCGGUGCU (Versión 1; SEQ ID NO: 5);

GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCA  
ACUUGAAAAAGUGGCACCGAGUCGGUGC (Versión 2; SEQ ID NO: 6);

45 GUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCAACUUGA  
AAAAGUGGCACCGAGUCGGUGC (Versión 3; SEQ ID NO: 7);

y

GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAUAAGGCUAGUCCGUU  
AUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (Versión 4; SEQ ID NO: 8);

50 Otras secuencias de andamios ilustrativas comprenden, consisten esencialmente en o consisten en:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCAACUUGA  
AAAAGUGGCACCGAGUCGGUGCUUUUUUUU (Versión 5; SEQ ID NO: 14);

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA  
AAAAGUGGCACCGAGUCGGUGCUUUU (versión 6; SEQ ID NO: 15);

o

GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUU  
AUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (versión 7; SEQ ID NO: 16);

5

Los ARN guía que se dirigen a cualquier secuencia diana de ARN guía pueden incluir, por ejemplo, un segmento de direccionamiento al ADN en el extremo 5' del ARN guía fusionado con cualquiera de las secuencias de andamio de ARN guía ilustrativas en el extremo 3' del ARN guía. Es decir, un segmento de direccionamiento al ADN puede unirse al extremo 5' de cualquiera de las SEQ ID NO: 5-8 para formar un ARN guía único (ARN guía quimérico). Asimismo, un segmento de direccionamiento al ADN puede unirse al extremo 5' de cualquiera de las SEQ ID NO: 14-16 para formar un ARN guía único (ARN guía quimérico). Las versiones de ARN guía 1, 2, 3 y 4, como se describe en otra parte de la presente descripción, se refieren a segmentos de direccionamiento al ADN (es decir, secuencias guía o guías) unidos con las versiones de andamio 1, 2, 3 y 4, respectivamente. Las versiones 5, 6 y 7 del ARN guía, como se describe en otra parte de la presente descripción, se refieren a segmentos de direccionamiento al ADN (es decir, secuencias guía o guías) unidos con las versiones 5, 6 y 7 del armazón, respectivamente.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada; direccionamiento subcelular; rastreo con una etiqueta fluorescente; un sitio de unión para una proteína o un complejo proteico, y similares). Los ejemplos de tales modificaciones incluyen, por ejemplo, un protector 5' (por ejemplo, un protector de 7-metilguanilato (m7G)); una cola poliadenilada 3' (es decir, una cola poli(A) 3'); una secuencia de riboconmutador (por ejemplo, para permitir la estabilidad regulada y/o la accesibilidad regulada por proteínas y/o complejos de proteínas); una secuencia de control de estabilidad; una secuencia que forma un dúplex de ARNs (es decir, una horquilla); una modificación o secuencia que dirige el ARN a una localización subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona el rastreo (por ejemplo, conjugación directa con una molécula fluorescente, conjugación con un resto 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, incluidas las ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares); y combinaciones de los mismos. Otros ejemplos de modificaciones incluyen estructuras dúplex de bucle de tallo manipuladas genéticamente, regiones salientes manipuladas genéticamente, horquillas 3' manipuladas genéticamente de la estructura dúplex de bucle de tallo o cualquier combinación de las mismas. Ver, por ejemplo, el documento US 2015/0376586. Un saliente puede ser una región desapareada de nucleótidos dentro del dúplex formado por la región similar a ARNcr y la región mínima similar a ARNtracr. Un saliente puede comprender, en un lado del dúplex, un 5'-XXXY-3' desapareado 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 nucleótido desapareada del otro lado del dúplex.

Los ácidos nucleicos no modificados pueden ser propensos a la degradación. Los ácidos nucleicos exógenos también pueden inducir una respuesta inmunitaria innata. Las modificaciones pueden ayudar a introducir estabilidad y reducir la inmunogenicidad. Los ARN guía pueden comprender nucleósidos modificados y nucleótidos modificados que incluyen, por ejemplo, uno o más de los siguientes: (1) alteración o reemplazo de uno o ambos de los oxígenos de fosfato no enlazantes y/o de uno o más de los oxígenos de fosfato enlazantes en el enlace de la cadena principal fosfodiéster; (2) alteración o reemplazo de un constituyente del azúcar ribosa tal como alteración o reemplazo del 2' hidroxilo en el azúcar ribosa; (3) reemplazo del resto fosfato con enlazadores defosfo; (4) modificación o reemplazo de una nucleobase de origen natural; (5) sustitución o modificación de la cadena principal de ribosa-fosfato; (6) modificación del extremo 3' o 5' del oligonucleótido (por ejemplo, eliminación, modificación o reemplazo de un grupo fosfato terminal o conjugación de un resto); y (7) modificación del azúcar. Otras posibles modificaciones del ARN guía incluyen modificaciones o sustitución de trectos de uracilo o poliuracilo. Ver, por ejemplo, los documentos WO 2015/048577 y US 2016/0237455. Pueden hacerse modificaciones similares a los ácidos nucleicos que codifican Cas, tales como los ARNm de Cas.

Como ejemplo, los nucleótidos en el extremo 5' o 3' de un ARN guía pueden incluir enlaces fosforotioato (por ejemplo, las bases pueden tener un grupo fosfato modificado que es un grupo fosforotioato). Por ejemplo, un ARN guía puede incluir enlaces de fosforotioato entre los 2, 3 o 4 nucleótidos terminales en el extremo 5' o 3' del ARN guía. Como otro ejemplo, los nucleótidos en el extremo 5' y/o 3' de un ARN guía pueden tener modificaciones 2'-O-metilo. Por ejemplo, un ARN guía puede incluir modificaciones 2'-O-metilo en los nucleótidos terminales 2, 3 o 4 en el extremo 5' y/o 3' del ARN guía (por ejemplo, el extremo 5'). Ver, por ejemplo, el documento WO 2017/173054 A1 y Finn y otros (2018) Cell Rep. 22(9):2227-2235. En un ejemplo específico, el ARN guía comprende análogos de 2'-O-metilo y enlaces internucleotídicos de fosforotioato 3' en los primeros tres residuos de ARN terminales 5' y 3'. En otro ejemplo específico, el ARN guía se modifica de manera que todos los grupos 2'OH que no interactúan con la

60

proteína Cas9 se reemplazan con análogos de 2'-O-metilo, y la región de la cola del ARN guía, que tiene una interacción mínima con Cas9, se modifica con enlaces internucleotídicos de fosforotioato en 5' y 3'. *Ver, por ejemplo*, Yin y otros (2017) Nat. Biotech. 35(12):1179-1187. Se proporcionan otros ejemplos de ARN guía modificados, por ejemplo, en el documento WO 2018/107028 A1.

Los ARNg pueden prepararse mediante varios otros métodos. Por ejemplo, los ARNg pueden prepararse mediante transcripción *in vitro* mediante el uso, por ejemplo, de ARN polimerasa T7 (*ver, por ejemplo*, los documentos WO 2014/089290 y WO 2014/065596). Los ARN guía también pueden ser una molécula producida sintéticamente preparada mediante síntesis química. Por ejemplo, un ARN guía puede sintetizarse químicamente para incluir análogos de 2'-O-metilo y enlaces internucleotídicos de fosforotioato 3' en los tres primeros residuos de ARN terminales 5' y 3'.

**Secuencias diana de ARN guía.** Los ADN diana para los ARN guía incluyen secuencias de ácido nucleico presentes en un ADN al que se unirá un segmento de direccionamiento al ADN de un ARNg, siempre que existan las condiciones suficientes para la unión. Las condiciones de unión de ADN/ARN adecuadas incluyen condiciones fisiológicas normalmente presentes en una célula. En la técnica se conocen otras condiciones de unión de ADN/ARN adecuadas (por ejemplo, condiciones en un sistema libre de células) (*ver, por ejemplo*, Molecular Cloning: A Laboratory Manual, 3ra Ed. (Sambrook y otros, Harbor Laboratory Press 2001). La cadena del ADN diana que es complementaria y se hibrida con el ARNg puede denominarse "cadena complementaria", y la cadena del ADN diana que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) puede llamarse "cadena no complementaria" o "cadena plantilla".

El ADN diana incluye tanto la secuencia de la cadena complementaria con la que se hibrida el ARN guía como la secuencia correspondiente de la cadena no complementaria (por ejemplo, adyacente al motivo adyacente de protoespaciador (PAM)). El término "secuencia diana de ARN guía" como se usa en la presente descripción se refiere específicamente a la secuencia en la cadena no complementaria correspondiente a (es decir, el complemento inverso de) la secuencia con la que se hibrida el ARN guía en la cadena complementaria. Es decir, la secuencia diana del ARN guía se refiere a la secuencia en la cadena no complementaria adyacente al PAM (por ejemplo, cadena arriba o 5' del PAM en el caso de Cas9). Una secuencia diana de ARN guía es equivalente al segmento de direccionamiento a ADN de un ARN guía, pero con timinas en lugar de uracilos. Como un ejemplo, una secuencia diana de ARN guía para una enzima SpCas9 puede referirse a la secuencia corriente arriba del PAM 5'-NGG-3' en la cadena no complementaria. Un ARN guía está diseñado para que tenga complementariedad con la cadena complementaria de un ADN diana, donde la hibridación entre el segmento de direccionamiento al ADN del ARN guía y la cadena complementaria del ADN diana promueve la formación de un complejo CRISPR. No se requiere necesariamente la complementariedad completa, siempre que haya suficiente complementariedad para provocar la hibridación y promover la formación de un complejo CRISPR. Si en la presente descripción se hace referencia a un ARN guía como que dirige una secuencia diana de ARN guía, lo que significa es que el ARN guía se hibrida con la secuencia de la cadena complementaria del ADN diana que es el complemento inverso de la secuencia diana del ARN guía en la cadena no complementaria.

Una secuencia de ADN diana o diana de ARN guía puede comprender cualquier polinucleótido y puede localizarse, por ejemplo, en el núcleo o citoplasma de una célula o dentro de un orgánulo de una célula, tal como una mitocondria o un cloroplasto. Una secuencia de ADN diana o diana de ARN guía puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia diana de ARN guía puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora) o puede incluir ambas.

La unión específica de sitio y la escisión de un ADN diana por una proteína Cas pueden producirse en localizaciones determinadas por (i) la complementariedad de apareamiento de bases entre el ARN guía y la cadena complementaria del ADN diana y (ii) un motivo corto, llamado el motivo adyacente de protoespaciador (PAM), en la cadena no complementaria del ADN diana. El PAM puede flanquear la secuencia diana del ARN guía. Opcionalmente, la secuencia diana de ARN guía puede estar flanqueada en el extremo 3' por el PAM (por ejemplo, para Cas9). Alternativamente, la secuencia diana de ARN guía puede estar flanqueada en el extremo 5' por el PAM (por ejemplo, para Cpf1). Por ejemplo, el sitio de escisión de las proteínas Cas puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) corriente arriba o corriente abajo de la secuencia PAM (por ejemplo, dentro de la secuencia diana de ARN guía). En el caso de SpCas9, la secuencia PAM (es decir, en la cadena no complementaria) puede ser 5'-N<sub>1</sub>GG-3', donde N<sub>1</sub> es cualquier nucleótido de ADN, y donde el PAM está inmediatamente 3' de la secuencia diana de ARN guía en la cadena no complementaria del ADN diana. Como tal, la secuencia correspondiente al PAM en la cadena complementaria (es decir, el complemento inverso) sería 5'-CCN<sub>2</sub>-3', donde N<sub>2</sub> es cualquier nucleótido de ADN y está inmediatamente 5' de la secuencia a la que pertenece el segmento de direccionamiento al ADN del ARN guía se hibrida en la cadena complementaria del ADN diana. En algunos de estos casos, N<sub>1</sub> y N<sub>2</sub> pueden ser complementarios y el par de bases N<sub>1</sub>-N<sub>2</sub> puede ser cualquier par de bases (por ejemplo, N<sub>1</sub>=C y N<sub>2</sub>=G; N<sub>1</sub>=G y N<sub>2</sub>=C; N<sub>1</sub>=A y N<sub>2</sub>=T; o N<sub>1</sub>=T, y N<sub>2</sub>=A). En el caso de Cas9 de *S. aureus*, el PAM puede ser NNGRRT o NNGRR, donde N puede ser A, G, C o T, y R puede ser G o A. En el caso de Cas9 de *C. jejuni*, el PAM puede ser, por ejemplo, NNNNACAC o NNNNRYAC, donde N puede ser A, G, C o T, y R puede ser G o A. En algunos casos (por

ejemplo, para FnCpf1), la secuencia de PAM puede estar corriente arriba del extremo 5' y tener la secuencia 5'-TTN-3'.

Un ejemplo de una secuencia diana de ARN guía es una secuencia de ADN de 20 nucleótidos que precede inmediatamente a un motivo NGG reconocido por una proteína SpCas9. Por ejemplo, dos ejemplos de secuencias diana de ARN guía más PAM son GN<sub>19</sub>NGG (SEQ ID NO: 9) o N<sub>20</sub>NGG (SEQ ID NO: 10). *Ver, por ejemplo*, el documento WO 2014/165825. La guanina en el extremo 5' puede facilitar la transcripción por la ARN polimerasa en las células. Otros ejemplos de secuencias diana de ARN guía más PAM pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN<sub>20</sub>NGG; SEQ ID NO: 11) para facilitar la transcripción eficaz por la polimerasa T7 *in vitro*. *Ver, por ejemplo*, el documento WO 2014/065596. Otras secuencias diana de ARN guía más PAM pueden tener una longitud entre 4-22 nucleótidos de SEQ ID NO: 9-11, incluidos la 5' G o GG y la 3' GG o NGG. Aún otras secuencias diana de ARN guía más PAM pueden tener una longitud entre 14 y 20 nucleótidos de SEQ ID NO: 9-11.

La formación de un complejo CRISPR hibridado con un ADN diana puede dar como resultado la escisión de una o ambas cadenas del ADN diana dentro o cerca de la región correspondiente a la secuencia diana de ARN guía (es decir, la secuencia diana de ARN guía en la cadena no complementaria del ADN diana y el complemento inverso en la cadena complementaria con la que se hibrida el ARN guía). Por ejemplo, el sitio de escisión puede estar dentro de la secuencia diana de ARN guía (por ejemplo, en una localización definida relativa a la secuencia de PAM). El "sitio de escisión" incluye la posición de un ADN diana en el que una proteína Cas produce una ruptura de cadena única o una ruptura de doble cadena. El sitio de escisión puede estar en una sola cadena (por ejemplo, cuando se usa una nickasa) o en ambas cadenas de un ADN de doble cadena. Los sitios de escisión pueden estar en la misma posición en ambas cadenas (produciendo extremos romos; por ejemplo, Cas9) o pueden estar en diferentes sitios en cada cadena (produciendo extremos escalonados (es decir, protuberancias); por ejemplo, Cpf1). Los extremos escalonados pueden producirse, por ejemplo, mediante el uso de dos proteínas Cas, cada una de las cuales produce una ruptura de cadena única en un sitio de escisión diferente en una cadena diferente, produciendo de este modo una ruptura de doble cadena. Por ejemplo, una primera nickasa puede crear una ruptura de cadena única en la primera cadena de ADN de doble cadena (ADNs), y una segunda nickasa puede crear una ruptura de cadena única en la segunda cadena de ADNs de modo que se creen secuencias protuberantes. En algunos casos, la secuencia diana de ARN guía o el sitio de escisión de la nickasa en la primera cadena está separada de la secuencia diana de ARN guía o el sitio de escisión de la nickasa en la segunda cadena por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1000 pares de bases.

### C. Otros agentes nucleasa

Cualquier otro tipo de agente nucleasa de corte raro conocido también puede usarse en los métodos descritos en la presente descripción. Un ejemplo de un agente nucleasa de este tipo es una nucleasa efectora de tipo activador de la transcripción (TALEN). Las nucleasas efectoras TAL son una clase de nucleasas específicas de secuencia que pueden usarse para generar rupturas de doble cadena en secuencias diana específicas en el ADN. Las nucleasas efectoras TAL se generan mediante la fusión de un efector similar al activador de la transcripción (TAL) nativo o manipulado genéticamente, o una parte funcional del mismo, con el dominio catalítico de una endonucleasa, tal como, por ejemplo, FokI. El dominio de unión al ADN de las efectoras TAL único y modular permite el diseño de proteínas potencialmente con cualquier especificidad de reconocimiento de ADN dada. Por lo tanto, los dominios de unión al ADN de las nucleasas efectoras TAL pueden manipularse genéticamente para reconocer sitios diana de ADN específicos y, por lo tanto, usarse para generar rupturas de doble cadena en las secuencias diana deseadas. *Ver el documento WO 2010/079430; Morbitzer y otros (2010) Proc. Natl. Acad. Sci. U.S.A. 107(50):21617-21622; Scholze & Boch (2010) Virulence 1:428-432; Christian y otros Genetics (2010) 186:757-761; Li y otros (2010) Nucleic Acids Res. (2010) 39(1):359-372; y Miller y otros (2011) Nat. Biotechnol. 29:143-148.*

Los ejemplos de nucleasas TAL adecuadas y de métodos para preparar nucleasas TAL adecuadas se describen, por ejemplo, en los documentos US 2011/0239315, US 2011/0269234, US 2011/0145940, US 2003/0232410, US 2005/0208489, US 2005/0026157, US 2005/0064474, US 2006/0188987 y US 2006/0063231.

En algunas TALEN, cada monómero de TALEN consiste en 33-35 repeticiones de TAL que reconocen un solo par de bases a través de dos residuos hipervariables. La TALEN puede ser una proteína quimérica que comprende un dominio de unión a ADN basado en repeticiones de TAL unido operativamente a una nucleasa independiente tal como una endonucleasa FokI. Por ejemplo, el agente nucleasa puede comprender un primer dominio de unión a ADN basado en repeticiones TAL y un segundo dominio de unión a ADN basado en repeticiones TAL, en donde cada uno de los dominios de unión a ADN basados en repeticiones TAL primero y segundo está unido operativamente a una nucleasa FokI, en donde el primero y el segundo dominio de unión a ADN basado en repeticiones TAL reconocen dos secuencias de ADN diana contiguas en cada cadena de la secuencia de ADN diana separadas por una secuencia espaciadora de longitud variable (12-20 pb), y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que hace que una doble cadena se rompa en una secuencia diana.

Otro ejemplo de un agente nucleasa adecuado es una nucleasa con dedos de zinc (ZFN). En algunos ZFN, cada monómero del ZFN comprende 3 o más dominios de unión a ADN basados en dedos de zinc, donde cada dominio

de unión a ADN basado en dedos de zinc se une a un subsitio de 3 pb. En otras ZFN, la ZFN es una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc unido operativamente a una nucleasa independiente tal como una endonucleasa FokI. Por ejemplo, el agente nucleasa puede comprender una primera ZFN y una segunda ZFN, en donde cada una de la primera ZFN y la segunda ZFN está unida operativamente a una subunidad de nucleasa FokI, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN diana contiguas en cada cadena de la secuencia de ADN diana separada por un espaciador de aproximadamente 5-7 pb, y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que genera una ruptura de doble cadena. Ver, por ejemplo, los documentos US 2006/0246567; US 2008/0182332; US 2002/0081614; US 2003/0021776; WO 2002/057308; US 2013/0123484; US 2010/0291048; WO 2011/017293; y Gaj y otros (2013) Trends Biotechnol., 31(7):397-405.

Otro tipo de agente nucleasa adecuado es una meganucleasa manipulada genéticamente. Las meganucleasas se han clasificado en cuatro familias basadas en motivos de secuencias conservadas, las familias son las familias LAGLIDADG, GIY-YIG, H-N-H y His-Cys box. Estos motivos participan en la coordinación de iones metálicos y la hidrólisis de enlaces fosodiéster. Las meganucleasas se destacan por sus largas secuencias diana y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Se conocen los dominios, la estructura y la función de la meganucleasa, ver, por ejemplo, Guhan y Muniyappa (2003) Crit. Rev. Biochem. Mol. Biol. 38:199-248; Lucas y otros (2001) Nucleic Acids Res. 29:960-9; Jurica y Stoddard, (1999) Cell. Mol. Life Sci. 55:1304-26; Stoddard (2006) Q. Rev. Biophys. 38:49-95; y Moure y otros (2002) Nat. Struct. Biol. 9:764. En algunos ejemplos, se usa una variante de origen natural y/o una meganucleasa derivada manipulada genéticamente. Se conocen métodos para modificar la cinética, las interacciones de los cofactores, la expresión, las condiciones óptimas y/o la especificidad de la secuencia diana y el tamizaje de la actividad. Ver, por ejemplo, Epinat y otros (2003) Nucleic Acids Res. 31:2952-62; Chevalier y otros (2002) Mol. Cell 10:895-905; Gimble y otros (2003) Mol. Biol. 334:993-1008; Seligman y otros (2002) Nucleic Acids Res. 30:3870-9; Sussman y otros (2004) J. Mol. Biol. 342:31-41; Rosen y otros (2006) Nucleic Acids Res. 34:4791-800; Chames y otros (2005) Nucleic Acids Res. 33:e178; Smith y otros (2006) Nucleic Acids Res. 34:e149; Gruen y otros (2002) Nucleic Acids Res. 30:e29; Chen y Zhao (2005) Nucleic Acids Res. 33:e154; WO 2005/105989; WO 2003/078619; WO 2006/097854; WO 2006/097853; WO 2006/097784; y WO 2004/031346.

Puede usarse cualquier meganucleasa, incluidas, por ejemplo, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIIVP, I-TiiI, I-PpoI, PI-Pspl, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-CsmI, I-CvuI, I-CvuAIP, I-DdII, I-DdIII, I-DiRI, I-DmOI, I-Hmul, I-Hmull, I-HsNIP, I-Llal, I-Msol, I-Naal, I-NanI, I-NciIP, I-NgrIP, I-NitI, I-Njal, I-Nsp236IP, I-PakI, I-PbolP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PoblP, I-PorI, I-PorIIP, I-PbplP, I-SpBetaIP, I-Scal, I-SexIP, I-SneiP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SqulP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, f-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIIP, PI-MtuHIIIP, PI-Pful, PI-Pfull, PI-PkolI, PI-PkolII, PI-Rma43812IP; PI-SpBetaIP, PI-SceI, PI-Tful, PI-Tfull, PI-Thyl, PI-TiII, PI-TiIII, o cualquiera de las variantes o fragmentos activas de los mismos.

Las meganucleasas pueden reconocer, por ejemplo, secuencias de ADN de doble cadena de 12 a 40 pares de bases. En algunos casos, la meganucleasa reconoce una secuencia diana perfectamente coincidente en el genoma. Algunas meganucleasas son nucleasas constitutivas. Un tipo de nucleasa constitutiva es una familia LAGLIDADG de nucleasas constitutivas que incluye, por ejemplo, I-SceI, I-CreI e I-DmOI.

#### VIII. Casetes de selección

Puede usarse cualquier casete de selección adecuado en los métodos descritos en la presente descripción. El término casete de selección se refiere a un casete de expresión que comprende una o más secuencias de control de la expresión (por ejemplo, un promotor para la expresión en una célula bacteriana y/u otras secuencias reguladoras, tales como potenciadores, elementos reguladores postranscripcionales y secuencias poli(A)) unidas operativamente a un ácido nucleico que codifica un marcador seleccionable. Los casetes de selección pueden permitir la selección en células bacterianas, o pueden permitir la selección tanto en células bacterianas como eucarióticas o de mamíferos. Como ejemplo, puede usarse un gen tal como neomicina fosfotransferasa. La neomicina fosfotransferasa confiere resistencia a kanamicina en células procariontes y resistencia a G418 en células eucariotas. Tal gen puede usarse, por ejemplo, en combinación con un sistema de promotor doble que combina un promotor eucariota (por ejemplo, un promotor de fosfoglicerato quinasa (PGK) eucariota) y un promotor procarionte (por ejemplo, un promotor EM7 procarionte).

Algunos casetes de selección que pueden usarse en los métodos descritos en la presente descripción pueden aportar resistencia a un antibiótico que, de otro modo, destruiría o inhibiría el crecimiento de las células bacterianas. Por ejemplo, un casete de selección puede aportar resistencia a kanamicina, espectinomicina, estreptomycin, ampicilina, carbenicilina, bleomicina, eritromicina, polimixina B, tetraciclina o cloranfenicol. Tales casetes de selección y genes que aportan resistencia a estos antibióticos y otros son bien conocidos. Las células que comprenden los casetes de selección pueden seleccionarse mediante el tratamiento de las células con el antibiótico. Aquellas células que son resistentes al antibiótico constituyen el casete de selección.

Otros casetes de selección pueden comprender genes indicadores que pueden usarse para seleccionar células que comprenden una modificación deseada. El término gen indicador se refiere a un ácido nucleico que tiene una secuencia que codifica un producto génico (típicamente una enzima) que se analiza de manera fácil y cuantificable cuando un constructo que comprende la secuencia del gen indicador unida operativamente a un elemento potenciador y/o promotor heterólogo se introduce en células que contienen (o que puede hacerse que contengan) los factores necesarios para la activación de los elementos promotores y/o potenciadores. Los ejemplos de genes indicadores incluyen, pero no se limitan a, genes que codifican proteínas fluorescentes. Una proteína indicadora se refiere a una proteína codificada por un gen indicador.

Una proteína indicadora fluorescente es una proteína indicadora que es detectable basado en la fluorescencia en donde la fluorescencia puede provenir directamente de la proteína indicadora, la actividad de la proteína indicadora en un sustrato fluorogénico, o una proteína con afinidad para unirse a un compuesto etiquetado con fluorescencia. Los ejemplos de proteínas fluorescentes incluyen proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP y ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrine, Venus, YPet, PhiYFP y ZsYellow1), proteínas fluorescentes azules (por ejemplo, BFP, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire y T-sapphire), proteínas fluorescentes cian (por ejemplo, CFP, eCFP, Cerulean, CyPet, AmCyan1 y Midoriishi-Cyan), proteínas fluorescentes rojas (por ejemplo, RFP, mKate, mKate2, mPlum, monómero DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tándem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry y Jred), proteínas fluorescentes naranjas (por ejemplo, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine y tdTomato) y cualquier otra proteína fluorescente adecuada cuya presencia en las células pueda detectarse mediante métodos de citometría de flujo. Las células que comprenden un gen indicador fluorescente pueden seleccionarse, por ejemplo, clasificando las células que comprenden la proteína indicadora fluorescente codificada por el gen.

#### IX. Modificaciones dirigidas

Pueden introducirse varios tipos de modificaciones genéticas dirigidas mediante el uso de los métodos descritos en la presente descripción. Tales modificaciones genéticas dirigidas pueden incluir, por ejemplo, la inserción de uno o más nucleótidos, la eliminación de uno o más nucleótidos o la sustitución (reemplazo) de uno o más nucleótidos. Tales inserciones, eliminaciones o reemplazos pueden dar como resultado, por ejemplo, una mutación puntual, una supresión de una secuencia de ácido nucleico de interés o una parte de la misma, una inserción de una secuencia de ácido nucleico de interés o una parte de la misma, un reemplazo de una secuencia de ácido nucleico endógena con una secuencia de ácido nucleico heteróloga o exógena, un reemplazo de una secuencia de ácido nucleico endógena con una secuencia de ácido nucleico homóloga u ortóloga (por ejemplo, intercambio de dominio, intercambio de exón, intercambio de intrón, intercambio de secuencia reguladora o intercambio de gen), alteración de un elemento regulador (por ejemplo, promotor o potenciador), una mutación de sentido erróneo, una mutación sin sentido, una mutación de cambio de marco, una mutación de truncamiento, una mutación nula o una combinación de las mismas. Por ejemplo, pueden cambiarse al menos 1, 2, 3, 4, 5, 7, 8, 9, 10 o más nucleótidos (por ejemplo, eliminar, insertar o sustituir) para formar la modificación genética dirigida. Las eliminaciones, inserciones o reemplazos pueden ser de cualquier tamaño, como se describe en otra parte de la presente descripción. Ver, por ejemplo, Wang y otros (2013) *Cell* 153:910-918; Mandalos y otros (2012) *PLOS One* 7:e45768; y Wang y otros (2013) *Nat Biotechnol.* 31:530-532.

Las eliminaciones, inserciones o reemplazos pueden tener cualquier longitud. El ácido nucleico eliminado, insertado o reemplazado puede tener, por ejemplo, de aproximadamente 1 pb a aproximadamente 5 pb, de aproximadamente 5 pb a aproximadamente 10 pb, de aproximadamente 10 pb a aproximadamente 50 pb, de aproximadamente 50 pb a aproximadamente 100 pb, de aproximadamente 100 pb a aproximadamente 200 pb, de aproximadamente 200 pb a aproximadamente 300 pb, de aproximadamente 300 pb a aproximadamente 400 pb, de aproximadamente 400 pb a aproximadamente 500 pb, de aproximadamente 500 pb a aproximadamente 1 kb, de aproximadamente 1 kb a aproximadamente 5 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, o de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, o de aproximadamente 400 kb a aproximadamente 500 kb.

Si se asocian diferentes versiones de una secuencia con un número de acceso en diferentes momentos, se entiende la versión asociada con el número de acceso en la fecha de presentación efectiva de esta solicitud. La fecha de presentación efectiva se refiere a la fecha de presentación anterior de la fecha de presentación concreta o la fecha de presentación de una solicitud de prioridad que se refiere al número de acceso, si es aplicable. Igualmente, si se publican diferentes versiones de una publicación, sitio web o similares en diferentes momentos, se entiende la versión más recientemente publicada en la fecha de presentación efectiva de la solicitud, a menos que se indique de cualquier otra manera.

## Breve descripción de las secuencias

Las secuencias de nucleótidos y aminoácidos enumeradas en el listado de secuencias adjunto se muestran mediante el uso de abreviaturas de letras estándar para las bases de nucleótidos y un código de tres letras para los aminoácidos. Las secuencias de nucleótidos siguen la convención estándar de comenzar en el extremo 5' de la secuencia y avanzar (es decir, de izquierda a derecha en cada línea) hasta el extremo 3'. Solo se muestra una cadena de cada secuencia de nucleótidos, pero se entiende que la cadena complementaria está incluida por cualquier referencia a la cadena mostrada. Cuando se proporciona una secuencia de nucleótidos que codifica una secuencia de aminoácidos, se entiende que también se proporcionan variantes de codones degenerados de la misma que codifican la misma secuencia de aminoácidos. Las secuencias de aminoácidos siguen la convención estándar de comenzar en el extremo amino de la secuencia y avanzar (es decir, de izquierda a derecha en cada línea) hasta el extremo carboxi.

Tabla 2. Descripción de Secuencias.

SEQ ID NO	Tipo	Descripción
1	Proteína	Proteína Cas9
2	ADN	ADN de Cas9
3	ARN	cola de ARNcr
4	ARN	ARNtracr
5	ARN	andamio de ARNg v1
6	ARN	andamio de ARNg v2
7	ARN	andamio de ARNg v3
8	ARN	andamio de ARNg v4
9	ADN	secuencia diana de ARN guía más PAM v1
10	ADN	secuencia diana de ARN guía más PAM v2
11	ADN	secuencia diana de ARN guía más PAM v3
12	ARN	ARNtracr v2
13	ARN	ARNtracr v3
14	ARN	andamio de ARNg v5
15	ARN	andamio de ARNg v6
16	ARN	andamio de ARNg v7

## Ejemplos

Ejemplo 1. Introducción sin cicatrices de una modificación dirigida en un vector de direccionamiento grande a través de la recombinación homóloga bacteriana y el ensamblaje de Gibson intramolecular

La tecnología de ensamblaje de Gibson une segmentos de ADN con extremos homólogos en una molécula única. Se diferencia de los ligamientos tradicionales entre extremos complementarios escalonados creados por enzimas de restricción en que puede usarse cualquier secuencia complementaria de un tamaño mínimo. Dado que la clonación a través de sitios de restricción generalmente da como resultado la incorporación de cicatrices de ADN exógenas (sitios de reconocimiento de enzimas) en el producto final, el ensamblaje de Gibson es ventajoso porque puede ser perfecto.

La reacción de ensamblaje de Gibson es isotérmica e implica tres enzimas diferentes: exonucleasa T5, ADN polimerasa y ligasa. Ver, por ejemplo, los documentos US 2010/0035768, US 2015/0376628, WO 2015/200334, y Gibson y otros (2009) Nat. Methods 6(5):343-345. La reacción comienza con la generación de extremos de ADN de cadena única debido a la actividad exonucleasa 5' a 3' por parte de la exonucleasa T5. Los fragmentos de ADN con extremos de cadena únicas complementarias se alinean a continuación mediante reglas simples de apareamiento de bases, y la ADN polimerasa llena los huecos moviéndose de 5' a 3'. La ligasa de ADN sella la mella final y el resultado es una molécula de ADN de doble cadena perfecta. Se ha demostrado que los extremos complementarios de cuarenta pares de bases son efectivos, y la secuencia real generalmente no es importante. Los fragmentos de partida pueden generarse mediante PCR, restricción o síntesis directa.

La construcción perfecta de ADN es de particular importancia cuando se crean líneas de animales transgénicos, ya que las cicatrices producidas por los sitios de restricción u otras manipulaciones pueden afectar negativamente la expresión génica si se producen en una región importante para la regulación. El direccionamiento al genoma de los mamíferos frecuentemente requiere la construcción de grandes vectores de direccionamiento con brazos de ADN largos para dirigir la recombinación homóloga, así como también casetes de resistencia a antibióticos para la selección de clones de células madre embrionarias. Los clones dirigidos correctamente contienen frecuentemente múltiples cicatrices necesarias para la construcción del vector, sin mencionar el casete de resistencia en sí. Para la ablación genética, estas lesiones pueden no ser importantes para el resultado final (un alelo nulo), pero siempre existe la posibilidad de que la expresión de los genes vecinos se vea afectada negativamente. Para modificaciones que no sean supresión, tal como inserción (por ejemplo, indicadores o alelos mutantes), la expresión fiel del locus dirigido suele ser importante para los estudios en cuestión. El ensamblaje de Gibson puede anular la necesidad de

algunas de estas cicatrices e incluso facilitar la construcción del propio vector en algunos casos, pero los sitios de restricción únicos pueden ser difíciles de encontrar.

La humanización, el reemplazo directo de un gen de ratón con su homólogo humano, requiere en particular uniones perfectas entre la secuencia de ratón y la humana para que la maquinaria de transcripción del ratón replique fielmente la expresión del nuevo alelo. Debe tener cuidado de ocultar las cicatrices de construcción y el casete de selección en regiones no codificantes que no afecten la regulación génica. A medida que los modelos animales se vuelven más complejos, pueden añadirse más modificaciones a las existentes, tales como mutaciones que causan enfermedades humanas en alelos humanizados. Los cambios adicionales pueden añadir aún más cicatrices y otro casete de selección a un locus de ratón ya altamente manipulado genéticamente, lo que aumenta la probabilidad de que la expresión se altere y el modelo de ratón no sea fiel a la enfermedad humana. Desde el punto de vista de la construcción, añadir un nuevo casete a un vector que ya contiene uno puede resultar complicado debido a la recombinación no deseada entre elementos de casete compartidos tales como promotores y señales poli(A), incluso si los dos casetes codifican selecciones diferentes.

En vista de estos obstáculos, hemos desarrollado métodos para simplificar la generación de vectores de direccionamiento que portan múltiples cambios tal como un alelo humanizado y una mutación de enfermedad superpuesta. Estos métodos permiten una construcción más sencilla y minimizan las cicatrices incorporadas en el modelo animal final.

En un primer método, se sintetiza una pequeña porción de ADN que porta una mutación deseada flanqueada por brazos de homología cortos (<500 pb). Unos pocos pares de bases corriente abajo de la mutación deseada, se selecciona y duplica una región de 40-50 pares de bases para crear repeticiones directas para flanquear sitios de restricción raros o secuencias diana de ARN guía Cas9 que flanquean un casete de resistencia. Esta pequeña construcción se recombina homológamente a continuación con un vector de direccionamiento de ratón establecido (tal como un vector de direccionamiento de humanización, con su propio casete de resistencia) mediante tecnología de recombinación. Después de confirmar que se ha incorporado la mutación deseada, el nuevo vector se corta con la guía Cas9/cortador raro, dejando caer el casete y exponiendo las repeticiones directas de 40-50 pares de bases. El ensamblaje de Gibson a continuación sella perfectamente la ruptura en una reacción intramolecular. El vector de direccionamiento resultante ahora porta la mutación deseada y no tiene cicatrices ni casetes adicionales además de los presentes originalmente en la humanización.

En un ejemplo específico, generamos un alelo que incorpora una mutación puntual en un constructo de direccionamiento (un gran vector de direccionamiento) que comprende un gen diana 1 humanizado. Ver la **Figura 1**. El objetivo era crear eficiente y perfectamente un gran vector de direccionamiento que comprendiera un gen diana 1 humanizado con la mutación puntual en lugar de tener que redirigir las células madre embrionarias (ES) de ratón humanizadas que comprendan el gen diana 1 humanizado para introducir la mutación puntual en esas células ES de ratón humanizadas. El constructo de direccionamiento inicial contenía la secuencia genómica del gen diana 1 humano desde el codón de inicio hasta el codón de finalización, incluidos todos los intrones, para reemplazar la secuencia genómica de ratón del gen diana 1 de ratón correspondiente desde el codón de inicio hasta el codón de finalización. Además, el inserto de ácido nucleico en la construcción de direccionamiento comprendía un casete de resistencia a higromicina de autoeliminación corriente abajo de la secuencia poli(A). Este vector de humanización inicial se modificó a continuación como se describió anteriormente con la mutación puntual y un casete de resistencia a neomicina, flanqueado por sitios de restricción *AscI* y repeticiones directas de 50 pares de bases de la secuencia del gen diana 1 humano justo corriente abajo de la mutación puntual. A continuación, se sintetizó un ácido nucleico para que comprenda un casete de neomicina EM7 flanqueado por sitios de restricción raros (*AscI*) y una repetición directa de 50 pb del exón en el que se iba a introducir la mutación, así como también cajas de homología corriente arriba y corriente abajo, incluida la mutación a introducir en la caja de homología corriente arriba. Ver la **Figura 2**. El casete resistente a neomicina se insertó en el medio del exón que se iba a mutar, pero debido a que el método es perfecto, el exón se recapituló al final del método. El ácido nucleico se linealizó mediante escisión con *HindIII*, y se usó la recombinación homóloga bacteriana para insertar el ácido nucleico sintético linealizado en el vector de direccionamiento grande que comprende el gen diana 1 humanizado. Ver, por ejemplo, el documento US 2004/0018626 y Valenzuela y otros (2003) *Nat. Biotechnol.* 21(6):652-659. El casete de neomicina se escindió con *AscI*, que eliminó el casete de neomicina y expuso las repeticiones directas. A continuación, la construcción se volvió a sellar mediante un ensamblaje de Gibson intramolecular, que resolvió las repeticiones directas en una sola copia, dejando el exón (que ahora comprende la mutación) intacto sin cicatrices. Después del ensamblaje de Gibson, la reacción se dirigió de nuevo con *AscI*, para cortar cualquier cosa que no eliminara los sitios de *AscI* durante el ensamblaje de Gibson, reduciendo de este modo el fondo. La secuenciación final confirmó la presencia de la mutación puntual y ningún cambio adicional con respecto al vector de direccionamiento original. El vector recién modificado se sometió a electroporación en células madre embrionarias de ratón y los clones positivos se identificaron mediante TAQMAN seguido de secuenciación de Sanger para confirmar la incorporación de la mutación puntual.

Ejemplo 2. Introducción sin cicatrices de una modificación dirigida en un vector de direccionamiento grande a través de la recombinación homóloga bacteriana y el ensamblaje de Gibson intermolecular

En un segundo método, se introduce una mutación deseada en un ADN de cromosoma artificial bacteriano (BAC) en dos etapas generales. En la primera etapa, la región de interés en el BAC (una región que abarca aproximadamente 100-200 pb en cada lado de la mutación) se elimina mediante recombinación homóloga bacteriana mediante el uso de un casete de selección flanqueado por un sitio de enzima de restricción de corte raro en cada lado. En la segunda etapa, la eliminación de BAC se reemplazan con la secuencia mutada deseada mediante el ensamblaje de Gibson mediante el uso de un fragmento de ADN de aproximadamente 200-500 pb que tiene extremos 5' y 3' heterólogos homólogos a la secuencia de BAC dirigida adyacente al sitio de corte raro. Para este propósito, el BAC dirigida en la primera etapa se digiere con la enzima de corte raro, exponiendo los dos extremos homólogos al fragmento mutado. La enzima de restricción también mantiene abierto el BAC dirigido, lo que permite una reacción de fondo baja sin necesidad de añadir un marcador de selección. Ver la **Figura 3**. Este método es particularmente beneficioso, por ejemplo, cuando es necesario insertar en un constructo fragmentos más grandes que no pueden obtenerse mediante PCR (por ejemplo, 15 kb o 30 kb). Por ejemplo, un fragmento tan grande puede cortarse de su origen como un BAC (por ejemplo, mediante el uso de CRISPR/Cas9) y a continuación insertarlo mediante el uso del ensamblaje de Gibson en un BAC modificado que porta homología a los extremos 5' y 3' de este fragmento, lo que crea de este modo el constructo de direccionamiento final.

En un ejemplo específico, este método se usó para introducir una mutación de corte y empalme en un gran vector de direccionamiento que comprende un gen diana 2 humanizado. El constructo de direccionamiento inicial contenía la secuencia genómica del gen diana 2 de tipo silvestre, incluidos los intrones, diseñada para reemplazar la secuencia genómica correspondiente del gen diana 2 de ratón desde el codón de inicio hasta antes del último exón, incluida la adición de un casete de resistencia a neomicina de autoeliminación en un intrón. Este vector de humanización inicial se modificó a continuación como se describe anteriormente con un casete de resistencia a higromicina, flanqueado por sitios de restricción *AscI* y una secuencia homóloga de cuarenta pares de bases corriente abajo y corriente arriba de la mutación de corte y empalme deseada. El casete de higromicina se cortó con *AscI*, y el constructo se volvió a sellar mediante ensamblaje de Gibson intermolecular con un fragmento de ADN que comprendía la mutación de corte y empalme flanqueada por extremos 5' y 3' heterólogos homólogos a la secuencia de constructo de direccionamiento dirigida adyacente al sitio de corte raro. La secuenciación final confirmó la presencia de la mutación de corte y empalme y ningún cambio adicional con respecto al vector de intermolecular original. El vector recién modificado se sometió a electroporación en células madre embrionarias de ratón y se identificó un clon positivo mediante TAQMAN, seguido de secuenciación de Sanger para confirmar la incorporación de la mutación de corte y empalme.

En un tercer método, se corta un fragmento de ADN humano de un cromosoma artificial bacteriano (BAC) mediante el uso de CRISPR/Cas9. Este fragmento de ADN humano se fusiona, mediante el ensamblaje de Gibson, con un BAC de ratón que se había dirigido previamente con un casete de selección. Se diseña un sitio de enzima de restricción de corte raro en la región donde se iba a integrar el fragmento humano. En el BAC de ratón dirigido, hay 40 pb de secuencias de homología en cada lado de este sitio de restricción de corte raro. Las secuencias de homología son homólogas a los extremos 5' y 3' del fragmento de ADN humano. El constructo final se selecciona en el mismo antibiótico al del BAC de ratón original dirigido. Aunque no se incorpora una nueva selección en la reacción final del ensamblaje de Gibson, se observa un fondo bajo. La adición de la enzima de restricción de corte raro después del ensamblaje de Gibson mantiene el fondo en un nivel bajo.

En un ejemplo específico, el experimento anterior consistía en incorporar un alelo que comprende la región del gen diana 3 que codifica el ectodominio de la proteína diana 3 en el gen diana 3 de ratón. El constructo de direccionamiento inicial contenía la secuencia genómica del gen diana 3 de ratón de tipo silvestre, incluidos los intrones. Se añadió un casete de resistencia a neomicina de autoeliminación mediante recombinación homóloga bacteriana, eliminando la región codificante del ectodominio del gen diana 3 de ratón. Corriente arriba del casete de resistencia a neomicina, había un sitio de restricción *SgrDI* que separa las regiones 5' y 3' de 40 pb de homología humana que interactuarán con el fragmento humano. Todas estas secuencias fueron incorporadas por recombinación homóloga bacteriana descrita previamente. Se cortó un fragmento de ADN humano de 32 kb de longitud de un BAC humano mediante CRISPR/Cas9, dejando los extremos 5' y 3' expuestos para la reacción de ensamblaje de Gibson intramolecular con el BAC dirigido de ratón que se abrió mediante digestión con *SgrDI*. El vector recién modificado se sometió a electroporación en células madre embrionarias de ratón y se identificó un clon positivo mediante TAQMAN.

Listado de secuencias

<110> Regeneron Pharmaceuticals, Inc.

5 <120> MÉTODOS PARA LA INTRODUCCIÓN SIN CICATRICES DE MODIFICACIONES DIRIGIDAS EN  
VECTORES DE DIRECCIONAMIENTO

<130> 057766-544999

10 <150> US 62/829,327  
<151> 2019-04-04

<160> 16

15 <170> PatentIn versión 3,5

<210> 1  
<211> 1370  
<212> PRT

20 <213> Secuencia Artificial

<220>  
<223> Sintético

25 <400> 1

# ES 2 923 629 T3

	Gly	Ala	Ala	Ser	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Asp	Ile	Gly	Thr
	1				5					10					15	
5	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser
				20					25					30		
10	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys
			35					40					45			
15	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala
		50					55					60				
20	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn
	65					70					75					80
25	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val
					85					90					95	
30	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu
				100					105					110		
35	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu
			115					120					125			
40	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys
		130					135					140				
45	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala
	145					150					155					160
50	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp
					165					170					175	
55	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val
				180					185					190		
60	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly
			195					200					205			
65	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg
		210					215					220				
70	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu
	225					230					235					240
75	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys
					245					250					255	
80	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp
				260					265						270	

ES 2 923 629 T3

Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln  
 275 280 285  
 5  
 Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu  
 290 295 300  
 10  
 Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu  
 305 310 315 320  
 15  
 Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr  
 325 330 335  
 20  
 Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu  
 340 345 350  
 25  
 Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly  
 355 360 365  
 30  
 Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu  
 370 375 380  
 35  
 Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp  
 385 390 395 400  
 40  
 Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln  
 405 410 415  
 45  
 Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe  
 420 425 430  
 50  
 Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr  
 435 440 445  
 55  
 Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg  
 450 455 460  
 60  
 Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn  
 465 470 475 480  
 65  
 Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu  
 485 490 495  
 Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro  
 500 505 510  
 Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr  
 515 520 525  
 Lys Val Lys Tyr Val Thr Glu Gly Met Lys Pro Ala Phe Leu Ser Gly  
 530 535 540

# ES 2 923 629 T3

5	Glu 545	Gln	Lys	Lys	Ala	Ile 550	Val	Asp	Leu	Leu	Phe 555	Lys	Thr	Asn	Arg	Lys 560
10	Val	Thr	Val	Lys 565	Gln	Leu	Lys	Glu	Asp	Tyr 570	Phe	Lys	Lys	Ile	Glu 575	Cys
15	Phe	Asp	Ser	Val 580	Glu	Ile	Ser	Gly	Val 585	Glu	Asp	Arg	Phe	Asn 590	Ala	Ser
20	Leu	Gly	Thr	Tyr 595	His	Asp	Leu	Leu	Lys 600	Ile	Ile	Lys	Asp 605	Lys	Asp	Phe
25	Leu 610	Asp	Asn	Glu	Glu	Asn	Glu 615	Asp	Ile	Leu	Glu	Asp 620	Ile	Val	Leu	Thr
30	Leu 625	Thr	Leu	Phe	Glu	Asp 630	Arg	Glu	Met	Ile	Glu 635	Glu	Arg	Leu	Lys	Thr 640
35	Tyr	Ala	His	Leu 645	Phe	Asp	Asp	Lys	Val	Met 650	Lys	Gln	Leu	Lys	Arg 655	Arg
40	Arg	Tyr	Thr	Gly 660	Trp	Gly	Arg	Leu	Ser 665	Arg	Lys	Leu	Ile	Asn 670	Gly	Ile
45	Arg	Asp	Lys 675	Gln	Ser	Gly	Lys	Thr 680	Ile	Leu	Asp	Phe	Leu 685	Lys	Ser	Asp
50	Gly 690	Phe	Ala	Asn	Arg	Asn	Phe 695	Met	Gln	Leu	Ile	His 700	Asp	Asp	Ser	Leu
55	Thr 705	Phe	Lys	Glu	Asp	Ile 710	Gln	Lys	Ala	Gln 715	Val	Ser	Gly	Gln	Gly	Asp 720
60	Ser	Leu	His	Glu 725	His	Ile	Ala	Asn	Leu	Ala 730	Gly	Ser	Pro	Ala	Ile 735	Lys
65	Lys	Gly	Ile	Leu 740	Gln	Thr	Val	Lys	Val 745	Val	Asp	Glu	Leu 750	Val	Lys	Val
70	Met	Gly	Arg 755	His	Lys	Pro	Glu	Asn 760	Ile	Val	Ile	Glu	Met 765	Ala	Arg	Glu
75	Asn	Gln	Thr 770	Thr	Gln	Lys	Gly 775	Gln	Lys	Asn	Ser	Arg 780	Glu	Arg	Met	Lys
80	Arg 785	Ile	Glu	Glu	Gly	Ile 790	Lys	Glu	Leu	Gly	Ser 795	Gln	Ile	Leu	Lys	Glu 800
85	His	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr

# ES 2 923 629 T3

	805	810	815
5	Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile 820 825 830		
10	Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe 835 840 845		
15	Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys 850 855 860		
20	Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys 865 870 875 880		
25	Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln 885 890 895		
30	Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu 900 905 910		
35	Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln 915 920 925		
40	Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys 930 935 940		
45	Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu 945 950 955 960		
50	Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys 965 970 975		
55	Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn 980 985 990		
60	Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser 995 1000 1005		
65	Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met 1010 1015 1020		
70	Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr 1025 1030 1035		
75	Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr 1040 1045 1050		
80	Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn 1055 1060 1065		

# ES 2 923 629 T3

	Gly	Glu	Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala
		1070					1075					1080			
5	Thr	Val	Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys
		1085					1090					1095			
10	Lys	Thr	Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu
		1100					1105					1110			
15	Pro	Lys	Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp
		1115					1120					1125			
20	Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr
		1130					1135					1140			
25	Ser	Val	Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys
		1145					1150					1155			
30	Leu	Lys	Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg
		1160					1165					1170			
35	Ser	Ser	Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly
		1175					1180					1185			
40	Tyr	Lys	Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr
		1190					1195					1200			
45	Ser	Leu	Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser
		1205					1210					1215			
50	Ala	Gly	Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys
		1220					1225					1230			
55	Tyr	Val	Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys
		1235					1240					1245			
60	Gly	Ser	Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln
		1250					1255					1260			
65	His	Lys	His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe
		1265					1270					1275			
70	Ser	Lys	Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu
		1280					1285					1290			
75	Ser	Ala	Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala
		1295					1300					1305			
80	Glu	Asn	Ile	Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro
		1310					1315					1320			

# ES 2 923 629 T3

Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr  
1325 1330 1335

5

Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser  
1340 1345 1350

10

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly  
1355 1360 1365

15

Gly Asp  
1370

5

<210> 2  
<211> 4104  
<212> ADN  
<213> Secuencia Artificial

<220>  
<223> Sintético

<400> 2

## ES 2 923 629 T3

	gataagaaat actcaatagg cttagatata ggcacaaata gcgtcggatg ggcggtgatc	60
	actgatgaat ataaggttcc gtctaaaaag ttcaaggttc tgggaaatac agaccgccac	120
5	agtatcaaaa aaaatcttat aggggctcct ttatttgaca gtggagagac agcggaaagcg	180
	actcgtctca aacggacagc tcgtagaagg tatacacgtc ggaagaatcg tatttgttat	240
10	ctacaggaga ttttttcaaa tgagatggcg aaagtagatg atagtttctt tcatcgactt	300
	gaagagtctt ttttggtgga agaagacaag aagcatgaac gtcacccctat ttttggaaat	360
	atagtagatg aagttgctta tcatgagaaa tatccaacta tctatcatct gcgaaaaaaaa	420
15	ttggtagatt ctactgataa agcggatttg cgcttaactt atttggcctt agcgcataatg	480
	attaagtttc gtggtcattt tttgattgag ggagatttaa atcctgataa tagtgatgtg	540
20	gacaaactat ttatccagtt ggtacaaacc tacaatcaat tatttgaaga aaaccctatt	600
	aacgcaagtg gagtagatgc taaagcgatt ctttctgcac gattgagtaa atcaagacga	660
	ttagaaaatc tcattgctca gctccccggt gagaagaaaa atggcttatt tgggaatctc	720
25	attgctttgt cattgggttt gaccctaata tttaaatcaa attttgattt ggcagaagat	780
	gctaaattac agctttcaaa agatacttac gatgatgatt tagataattt attggcgcaa	840
30	attggagatc aatatgctga tttgtttttg gcagctaaga atttatcaga tgctatttta	900
	ctttcagata tcctaagagt aaatactgaa ataactaagg ctcccctatc agcttcaatg	960
	attaaacgct acgatgaaca tcatcaagac ttgactcttt taaaagcttt agttcgacaa	1020
35	caacttcag aaaagtataa agaaatcttt tttgatcaat caaaaaacgg atatgcaggt	1080
	tatattgatg ggggagctag ccaagaagaa tttataaat ttatcaaacc aattttagaa	1140
40	aaaatggatg gtactgagga attattggtg aaactaaatc gtgaagattt gctgcgcaag	1200
	caacggacct ttgacaacgg ctctattccc catcaaatc acttgggtga gctgcatgct	1260
	attttgagaa gacaagaaga cttttatcca tttttaaag acaatcgtga gaagattgaa	1320
45	aaaatcttga cttttcgaat tccttattat gttggtccat tggcgcgtgg caatagtcgt	1380
	tttgcattga tgactcggaa gtctgaagaa acaattaacc catggaattt tgaagaagtt	1440
50	gtcgataaag gtgcttcagc tcaatcattt attgaacgca tgacaaactt tgataaaaat	1500
	cttccaaatg aaaaagtact accaaaacat agtttgcttt atgagtattt tacggtttat	1560
	aacgaattga caaaggtcaa atatgttact gaaggatgc gaaaaccagc atttctttca	1620
55	ggtgaacaga agaaagccat tgttgattta ctcttcaaaa caaatcgaaa agtaaccggt	1680
	aagcaattaa aagaagatta tttcaaaaa atagaatggt ttgatagtggt tgaatttca	1740
60	ggagttgaag atagatttaa tgcttcatta ggtacctacc atgatttgct aaaaattatt	1800
	aaagataaag attttttgga taatgaagaa aatgaagata tcttagagga tattgtttta	1860
	acattgacct tatttgaaga tagggagatg attgaggaaa gacttaaaac atatgctcac	1920
65	ctctttgatg ataagtgat gaaacagctt aaacgtcgcc gttatactgg ttggggacgt	1980
	ttgtctcgaa aattgattaa tggatttagg gataagcaat ctggcaaaac aatattagat	2040

## ES 2 923 629 T3

	tttttgaat cagatggttt tgccaatcgc aattttatgc agctgatcca tgatgatagt	2100
5	ttgacattta aagaagacat tcaaaaagca caagtgtctg gacaaggcga tagtttacat	2160
	gaacatattg caaatttagc tggtagccct gctattaaaa aaggtatttt acagactgta	2220
	aaagtgttg atgaattggc caaagtaatg gggcggcata agccagaaaa tatcgttatt	2280
10	gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg	2340
	aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctggt	2400
15	gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctcaaaa tggaagagac	2460
	atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt	2520
	gttccacaaa gtttccttaa agacgattca atagacaata aggtcttaac gcgttctgat	2580
20	aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac	2640
	tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg	2700
25	aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg	2760
	gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact	2820
	aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa	2880
30	ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac	2940
	catcatgcc atgatgcgta tctaaatgcc gtcgttgga ctgctttgat taagaaatat	3000
35	ccaaaacttg aatcggagtt tgtctatggc gattataaag tttatgatgt tcgtaaaatg	3060
	attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaata	3120
	atcatgaact tcttcaaaac agaaattaca cttgcaaattg gagagattcg caaacgcctt	3180
40	ctaactgaaa ctaatgggga aactggagaa attgtctggg ataaaggcg agattttgcc	3240
	acagtgcgca aagtattgtc catgcccga gtcaatattg tcaagaaaac agaagtacag	3300
45	acaggcggat tctcaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct	3360
	cgtaaaaaag actgggatcc aaaaaatat ggtggttttg atagtccaac ggtagcttat	3420
	tcagtccctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa	3480
50	gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt	3540
	ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat	3600
55	agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa	3660
	aaaggaaatg agctggctct gccaaagcaaa tatgtgaatt ttttatattt agctagtcat	3720
	tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattggt tgtggagcag	3780
60	cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt	3840
	ttagcagatg ccaatttaga taaagttctt agtgcataa acaaacatag agacaacca	3900
	atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc	3960
65	gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc taaaaagaa	4020

# ES 2 923 629 T3

	gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat	4080
	ttgagtcagc taggaggtga ctga	4104
5	<210> 3 <211> 16 <212> ARN <213> Secuencia Artificial	
10	<220> <223> Sintético	
15	<400> 3 guuuuagagc uaugcu	16
20	<210> 4 <211> 67 <212> ARN <213> Secuencia Artificial	
25	<220> <223> Sintético	
30	<400> 4 agcauagcaa guaaaauaa ggcuaguccg uuaucaacu gaaaaagugg caccgagucg	60
35	gugcuuu	67
40	<210> 5 <211> 77 <212> ARN <213> Secuencia Artificial	
45	<220> <223> Sintético	
50	<400> 5 guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguaaucaac uugaaaaagu	60
55	ggcaccgagu cggugcu	77
60	<210> 6 <211> 82 <212> ARN <213> Secuencia Artificial	
65	<220> <223> Sintético	
70	<400> 6 guuggaacca uucaaacag cauagcaagu uaaaauaagg cuaguccguu aucaacuuga	60
75	aaaaguggca ccgagucggu gc	82
80	<210> 7 <211> 76 <212> ARN <213> Secuencia Artificial	

# ES 2 923 629 T3

<220>  
 <223> Sintético

5 <400> 7

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaagu 60  
 ggcaccgagu cggugc 76

10

<210> 8  
 <211> 86  
 <212> ARN  
 <213> Secuencia Artificial

15

<220>  
 <223> Sintético

20 <400> 8

guuuuagagc uaugcuggaa acagcauagc aaguuaaaau aaggcuaguc cguuaucaac 60  
 uugaaaagu ggcaccgagu cggugc 86

25

<210> 9  
 <211> 23  
 <212> ADN  
 <213> Secuencia Artificial

30

<220>  
 <223> Sintético

35

<220>  
 <221> misc\_característica  
 <222> (2)..(21)  
 <223> n es a, c, g o t

40 <400> 9

gnnnnnnnnn nnnnnnnnnn ngg 23

45

<210> 10  
 <211> 23  
 <212> ADN  
 <213> Secuencia Artificial

50

<220>  
 <223> Sintético

55

<220>  
 <221> misc\_característica  
 <222> (1)..(21)  
 <223> n es a, c, g o t

60 <400> 10

nnnnnnnnnn nnnnnnnnnn ngg 23

65

<210> 11  
 <211> 25  
 <212> ADN  
 <213> Secuencia Artificial

# ES 2 923 629 T3

<220>  
 <223> Sintético

5

<220>  
 <221> misc\_característica  
 <222> (3)..(23)  
 <223> n es a, c, g o t

10

<400> 11  
 ggnnnnnnnn nnnnnnnnnn nnngg 25

15

<210> 12  
 <211> 72  
 <212> ARN  
 <213> Secuencia Artificial

20

<220>  
 <223> Sintético

<400> 12

25

aaacagcaua gcaaguuaaa auaaggcuag uccguuauca acuugaaaaa guggcaccga 60  
 gucggugcuu uu 72

30

<210> 13  
 <211> 82  
 <212> ARN  
 <213> Secuencia Artificial

35

<220>  
 <223> Sintético

<400> 13

40

guuggaacca uucaaacag cauagcaagu uaaaauaagg cuaguccguu aucaacuuga 60  
 aaaaguggca ccgagucggu gc 82

45

<210> 14  
 <211> 83  
 <212> ARN  
 <213> Secuencia Artificial

50

<220>  
 <223> Sintético

<400> 14

55

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60  
 ggcaccgagu cggugcuuuu uuu 83

60

<210> 15  
 <211> 80  
 <212> ARN  
 <213> Secuencia Artificial

65

<220>  
 <223> Sintético

# ES 2 923 629 T3

<400> 15

5 guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaagu 60  
ggcaccgagu cggugcuuuu 80

<210> 16  
<211> 92  
<212> ARN  
<213> Secuencia Artificial

15 <220>  
<223> Sintético

<400> 16

20 guuuuagagc uaugcuggaa acagcauagc aaguuaaaau aaggcuaguc cguuaucaac 60  
uugaaaagu ggcaccgagu cggugcuuuu uu 92

## REIVINDICACIONES

- 5 1. Un método para introducir una modificación genética dirigida sin cicatrices en un vector de direccionamiento preexistente, que comprende:
- 10 (a) realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de modificación en una población de células bacterianas, en donde el casete de modificación comprende la modificación genética dirigida y comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de direccionamiento preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente, en donde el inserto de ácido nucleico comprende de 5' a 3':
- 15 (i) una primera secuencia de repetición;
- (ii) un primer sitio diana para un primer agente nucleasa;
- (iii) un casete de selección;
- (iv) un segundo sitio diana para un segundo agente nucleasa; y
- (v) una segunda secuencia de repetición idéntica a la primera secuencia de repetición;
- 20 (b) seleccionar células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección;
- (c) escindir el primer sitio diana en el vector de direccionamiento modificado con el primer agente nucleasa y escindir el segundo sitio diana en el vector de direccionamiento modificado con el segundo agente nucleasa para eliminar el casete de selección y exponer la primera secuencia de repetición y la segunda secuencia de repetición en el vector de direccionamiento modificado; y
- 25 (d) ensamblar la primera secuencia de repetición expuesta con la segunda secuencia de repetición expuesta en una reacción de ensamblaje intramolecular *in vitro* para generar el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices,
- 30 en donde no están presentes ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa y solo está presente una copia única de la secuencia de repetición en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices.
- 35 2. El método de la reivindicación 1, en donde:
- (I)
- (a) la secuencia de repetición es idéntica a una secuencia en el vector de direccionamiento preexistente, o
- 40 (b) la modificación genética dirigida comprende una inserción, y la secuencia de repetición es idéntica al extremo 5' o al extremo 3' de la inserción; y/o
- (II) la secuencia de repetición tiene una longitud de al menos 20 nucleótidos, opcionalmente en donde la secuencia de repetición tiene una longitud entre 20 nucleótidos y 100 nucleótidos.
- 45 3. El método de la reivindicación 1 o 2, en donde:
- (I) el casete de modificación es un ácido nucleico lineal de doble cadena; y/o
- (II) el casete de modificación tiene una longitud de 1 kb a 15 kb.
- 50 4. El método de cualquier reivindicación anterior, en donde el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de al menos 35 nucleótidos, o en donde el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de entre 35 nucleótidos y 500 nucleótidos.
- 55 5. El método de cualquier reivindicación anterior, en donde el primer agente nucleasa y/o el segundo agente nucleasa es un agente nucleasa de corte raro.
6. El método de cualquier reivindicación anterior, en donde:
- 60 (I) el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente; y/o
- (II) el primer sitio diana es idéntico al segundo sitio diana, y el primer agente nucleasa es idéntico al segundo agente nucleasa.
- 65 7. El método de cualquier reivindicación anterior, en donde:

- (I) el primer agente nucleasa y/o el segundo agente nucleasa comprenden una enzima de restricción de corte raro, opcionalmente en donde la enzima de restricción de corte raro es NotI, XmaII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AclI, CciNI, FspAI, MauBI, MreI, MssI, PaeAI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmlI, SrfI, Sse2321, Sse83871, LguI, PciSI, AarI, AjuI, AluI, BarI, PpiI, o PspI; o
- (II) el primer agente nucleasa y/o el segundo agente nucleasa es una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (Cas) y un ARN guía (ARNg), una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa manipulada genéticamente,
- opcionalmente en donde el primer agente nucleasa y/o el segundo agente nucleasa es la proteína Cas y el ARNg, en donde la proteína Cas es una proteína Cas9, y en donde el ARNg comprende un ARN CRISPR (ARNcr) que dirige un ARN CRISPR transactivador (ARNtracr).
8. El método de cualquier reivindicación anterior, en donde la modificación genética dirigida comprende:
- (I) una modificación en el brazo de homología 5' o el brazo de homología 3', o
- (II) una modificación en el inserto de ácido nucleico.
9. El método de cualquier reivindicación anterior, en donde la modificación genética dirigida comprende una mutación puntual, una eliminación, una inserción, un reemplazo o una combinación de los mismos.
10. El método de cualquiera de las reivindicaciones anteriores, en donde el casete de selección aporta resistencia a un antibiótico,
- opcionalmente en donde el casete de selección aporta resistencia a ampicilina, cloranfenicol, tetraciclina, kanamicina, espectinomina, estreptomina, carbenicilina, bleomicina, eritromicina o polimixina B.
11. El método de cualquier reivindicación anterior, en donde el vector de direccionamiento preexistente es un vector de direccionamiento grande de una longitud de al menos 10 kb, o en donde el vector de direccionamiento preexistente tiene una longitud de al menos 100 kb.
12. El método de cualquier reivindicación anterior, en donde el vector de direccionamiento preexistente comprende un segundo casete de selección,
- opcionalmente en donde el segundo casete de selección aporta resistencia a un antibiótico,
- opcionalmente en donde cada uno del casete de selección en el casete de modificación y el segundo casete de selección en el vector de direccionamiento preexistente aporta resistencia a un antibiótico diferente, y/u
- opcionalmente en donde el segundo casete de selección permite la selección tanto en células bacterianas como de mamífero.
13. El método de cualquier reivindicación anterior, en donde la etapa (c) se produce *in vitro*.
14. El método de cualquier reivindicación anterior, en donde la etapa (d) comprende:
- (i) poner en contacto el vector de direccionamiento modificado con una exonucleasa para exponer secuencias complementarias entre la primera secuencia de repetición y la segunda secuencia de repetición;
- (ii) hibridar las secuencias complementarias expuestas;
- (iii) extender los extremos 3' de las secuencias complementarias hibridadas; y
- (iv) ligar la secuencia complementaria hibridada, y
- opcionalmente en donde la etapa (d) comprende incubar el vector de direccionamiento modificado con una exonucleasa, una ADN polimerasa y una ADN ligasa.
15. El método de cualquier reivindicación anterior, que comprende además:
- (e) tratar el vector de direccionamiento con el primer agente nucleasa y el segundo agente nucleasa después del ensamblaje *in vitro* en la etapa (d) para verificar que no estén presentes ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa.
16. Un método para introducir una modificación genética dirigida sin cicatrices en un vector de direccionamiento preexistente, que comprende:
- (a) realizar una recombinación homóloga bacteriana entre el vector de direccionamiento preexistente y un casete de eliminación en una población de células bacterianas,

en donde el casete de eliminación comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' correspondiente a una secuencia diana 5' en el vector de direccionamiento preexistente y un brazo de homología 3' correspondiente a una secuencia diana 3' en el vector preexistente, en donde la secuencia diana 5' y la secuencia diana 3' flanquean una región del vector de direccionamiento preexistente en la que se va a introducir la modificación genética dirigida, y en donde el inserto de ácido nucleico comprende de 5' a 3':

- (i) un primer sitio diana para un primer agente nucleasa;
- (ii) un casete de selección; y
- (iii) un segundo sitio diana para un segundo agente nucleasa;

(b) seleccionar células bacterianas que comprenden un vector de direccionamiento modificado que comprende el casete de selección;

(c) escindir el primer sitio diana en el vector de direccionamiento modificado con el primer agente nucleasa y escindir el segundo sitio diana en el vector de direccionamiento modificado con el segundo agente nucleasa para eliminar el casete de selección y exponer una secuencia final corriente arriba y una secuencia final corriente abajo en el vector de direccionamiento modificado; y

(d) ensamblar en una reacción de ensamblaje *in vitro* el vector de direccionamiento escindido con un casete de modificación que comprende la modificación genética dirigida flanqueada por una secuencia final corriente arriba que se solapa con la secuencia final corriente arriba en el vector de direccionamiento modificado y una secuencia final corriente abajo que se solapa con la secuencia final corriente abajo en el vector de direccionamiento modificado para generar el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices,

en donde ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa están presentes en el vector de direccionamiento que comprende la modificación genética dirigida sin cicatrices.

17. El método de la reivindicación 16, en donde:

- (I) el casete de eliminación tiene una longitud de 1 kb a 15 kb, y/o
- (II) el casete de eliminación es un ácido nucleico lineal de doble cadena.

18. El método de la reivindicación 16 o 17, en donde el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de al menos 35 nucleótidos, o en donde el brazo de homología 5' y el brazo de homología 3' tienen cada uno una longitud de entre 35 nucleótidos y 500 nucleótidos.

19. El método de cualquiera de las reivindicaciones 16-18, en donde el primer agente nucleasa y/o el segundo agente nucleasa es un agente nucleasa de corte raro.

20. El método de cualquiera de las reivindicaciones 16-19, en donde:

- (I) el primer sitio diana y/o el segundo sitio diana no están presentes en el vector de direccionamiento preexistente; y/o
- (II) el primer sitio diana es idéntico al segundo sitio diana, y el primer agente nucleasa es idéntico al segundo agente nucleasa.

21. El método de cualquiera de las reivindicaciones 16-20, en donde:

(I) el primer agente nucleasa y/o el segundo agente nucleasa comprende una enzima de restricción de corte raro, opcionalmente en donde la enzima de restricción de corte raro es NotI, XmaII, SstII, Sall, NruI, NheI, Nb.BbvCI, BbvCI, AscI, AsiSI, FseI, PaeI, PmeI, SbfI, SgrAI, SwaI, BspQI, SapI, SfiI, CspCI, AclI, CciNI, FspAI, MauBI, MreI, MssI, PaeAI, RgaI, RgiI, SdaI, SfaAI, SgfI, SgrDI, SgsI, SmlI, SrfI, Sse2321, Sse83871, LglI, PciSI, AarI, AjuI, AclI, BarI, PpiI, o PstI; o

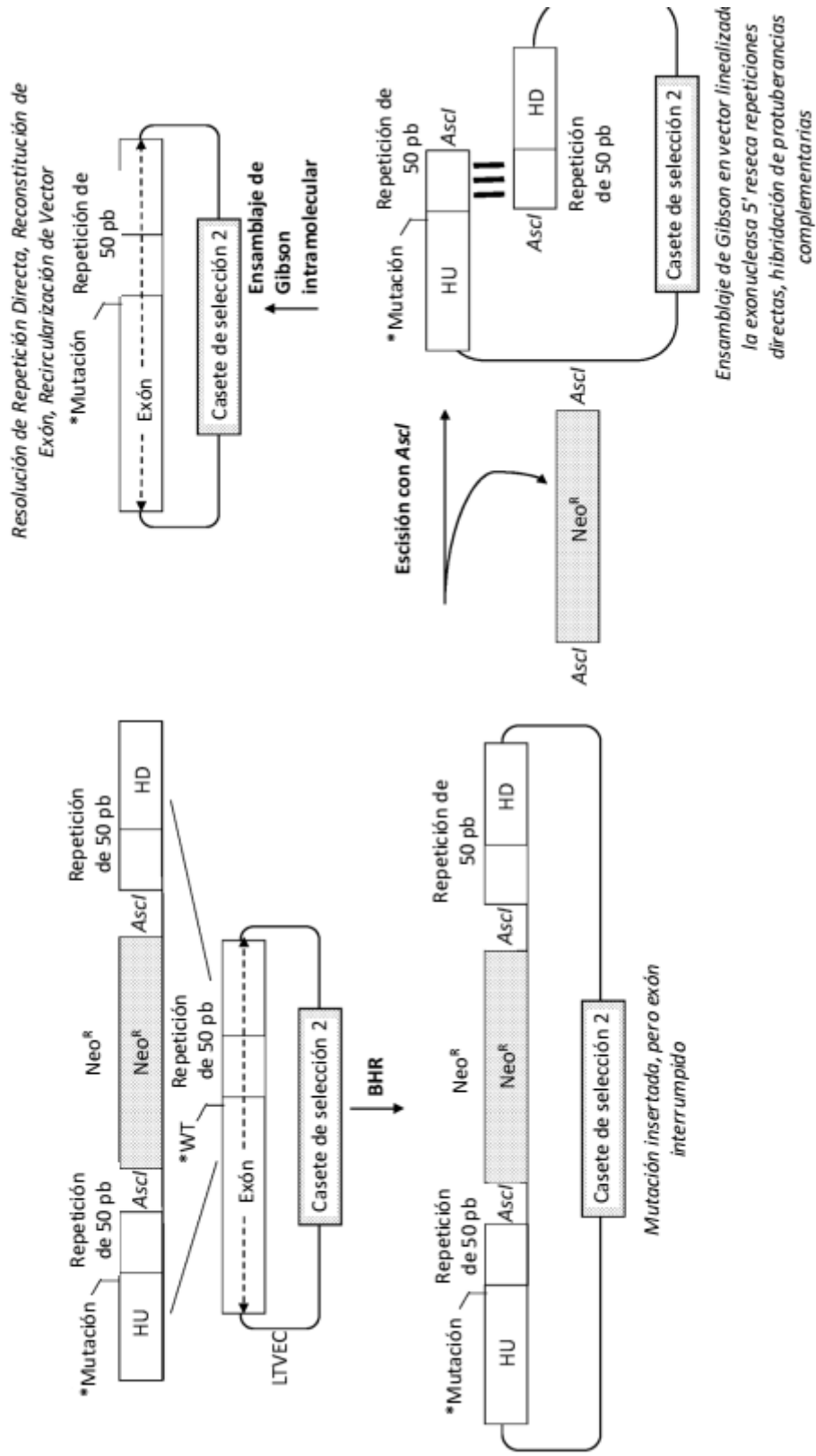
(II) el primer agente nucleasa y/o el segundo agente nucleasa es una proteína asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) (Cas) y un ARN guía (ARNg), una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa manipulada genéticamente,

opcionalmente en donde el primer agente nucleasa y/o el segundo agente nucleasa es la proteína Cas y el ARNg, en donde la proteína Cas es una proteína Cas9, y en donde el ARNg comprende un ARN CRISPR (ARNcr) que dirige un ARN CRISPR transactivador (ARNtracr).

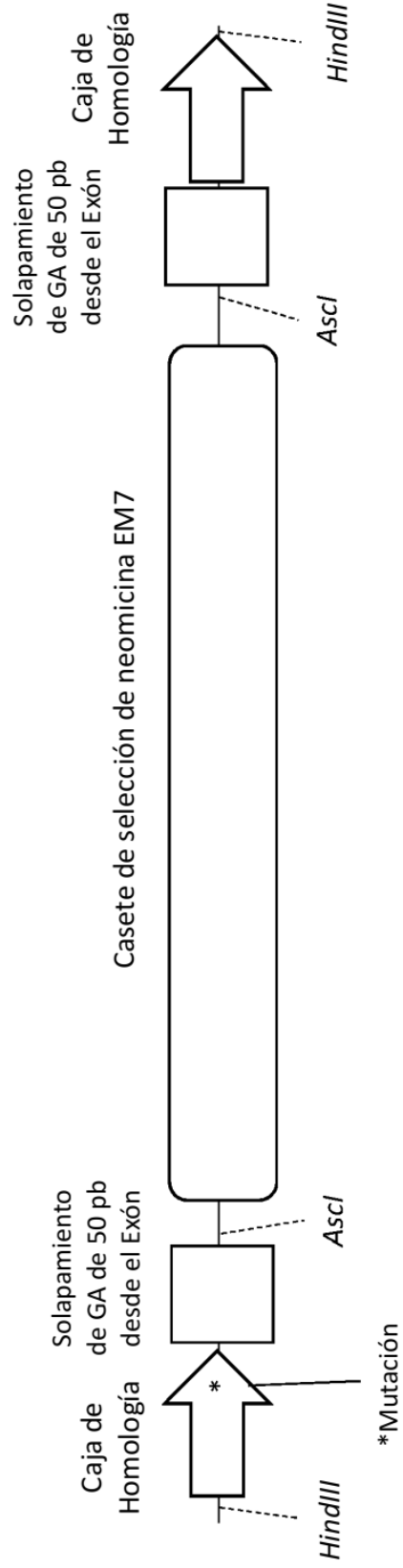
22. El método de cualquiera de las reivindicaciones 16-21, en donde el casete de selección aporta resistencia a un antibiótico, opcionalmente en donde el casete de selección aporta resistencia a ampicilina, cloranfenicol,

tetraciclina, kanamicina, espectinomycin, estreptomycin, carbenicilina, bleomicina, eritromicina o polimixina B.

- 5 23. El método de cualquiera de las reivindicaciones 16-22, en donde el vector de direccionamiento preexistente es un vector de direccionamiento grande de una longitud de al menos 10 kb, o en donde el vector de direccionamiento preexistente tiene una longitud de al menos 100 kb.
- 10 24. El método de cualquiera de las reivindicaciones 16-23, en donde el vector de direccionamiento preexistente comprende un segundo casete de selección,  
 opcionalmente en donde el segundo casete de selección aporta resistencia a un antibiótico,  
 opcionalmente en donde cada uno del casete de selección en el casete de eliminación y el segundo casete de selección en el vector de direccionamiento preexistente aporta resistencia a un antibiótico diferente, y/u  
 15 opcionalmente en donde el segundo casete de selección permite la selección tanto en células bacterianas como de mamíferos.
- 20 25. El método de cualquiera de las reivindicaciones 16-24, en donde la longitud de solapamiento entre la secuencia final corriente arriba en el casete de modificación y la secuencia final corriente arriba en el vector de direccionamiento modificado y/o la longitud de solapamiento entre la secuencia final corriente abajo en el casete de modificación y la secuencia final corriente abajo en el vector de direccionamiento modificado tiene una longitud de al menos 20 nucleótidos,  
 o en donde la longitud de solapamiento entre la secuencia final corriente arriba en el casete de modificación y la secuencia final corriente arriba en el vector de direccionamiento modificado y/o la longitud de solapamiento  
 25 entre la secuencia final corriente abajo en el casete de modificación y la secuencia final corriente abajo en el vector de direccionamiento modificado tiene una longitud de entre 20 y 100 nucleótidos.
- 30 26. El método de cualquiera de las reivindicaciones 16-25, en donde la etapa (c) se produce *in vitro*.
- 30 27. El método de cualquiera de las reivindicaciones 16-26, en donde la etapa (d) comprende:  
 (i) poner en contacto el vector de direccionamiento escindido y el casete de modificación con una exonucleasa para exponer secuencias complementarias entre las secuencias finales en el vector de  
 35 direccionamiento modificado y las secuencias finales en el casete de modificación;  
 (ii) hibridar las secuencias complementarias expuestas;  
 (iii) extender los extremos 3' de las secuencias complementarias hibridadas; y  
 (iv) ligar la secuencia complementaria hibridada, y  
 opcionalmente en donde la etapa (d) comprende incubar el vector de direccionamiento escindido y el casete de modificación con una exonucleasa, una ADN polimerasa y una ADN ligasa.
- 40 28. El método de cualquiera de las reivindicaciones 16-27, en donde:  
 (I) el casete de modificación es un ácido nucleico lineal de doble cadena;  
 45 (II) el casete de modificación tiene una longitud de al menos 200 nucleótidos, opcionalmente en donde el casete de modificación tiene una longitud de al menos 10 kb; y/o  
 (III) el casete de modificación tiene un tamaño que no puede sintetizarse o generarse directamente mediante la reacción en cadena de la polimerasa.
- 50 29. El método de cualquiera de las reivindicaciones 16-28, en donde la modificación genética dirigida comprende una mutación puntual, una eliminación, una inserción, un reemplazo o una combinación de los mismos.
- 55 30. El método de cualquiera de las reivindicaciones 16-29, que comprende además:  
 (e) tratar el vector de direccionamiento con el primer agente nucleasa y el segundo agente nucleasa después del ensamblaje *in vitro* en la etapa (d) para verificar que no estén presentes ni el primer sitio diana para el primer agente nucleasa ni el segundo sitio diana para el segundo agente nucleasa.

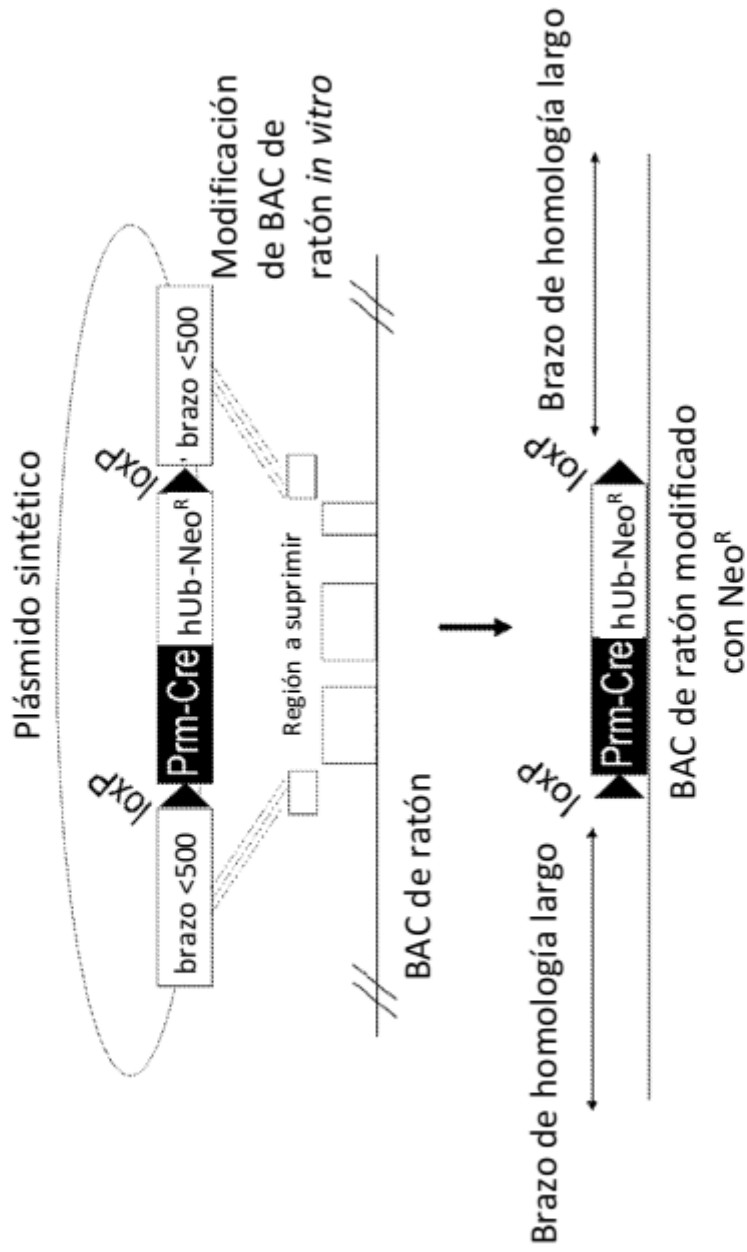


**FIGURA 7**

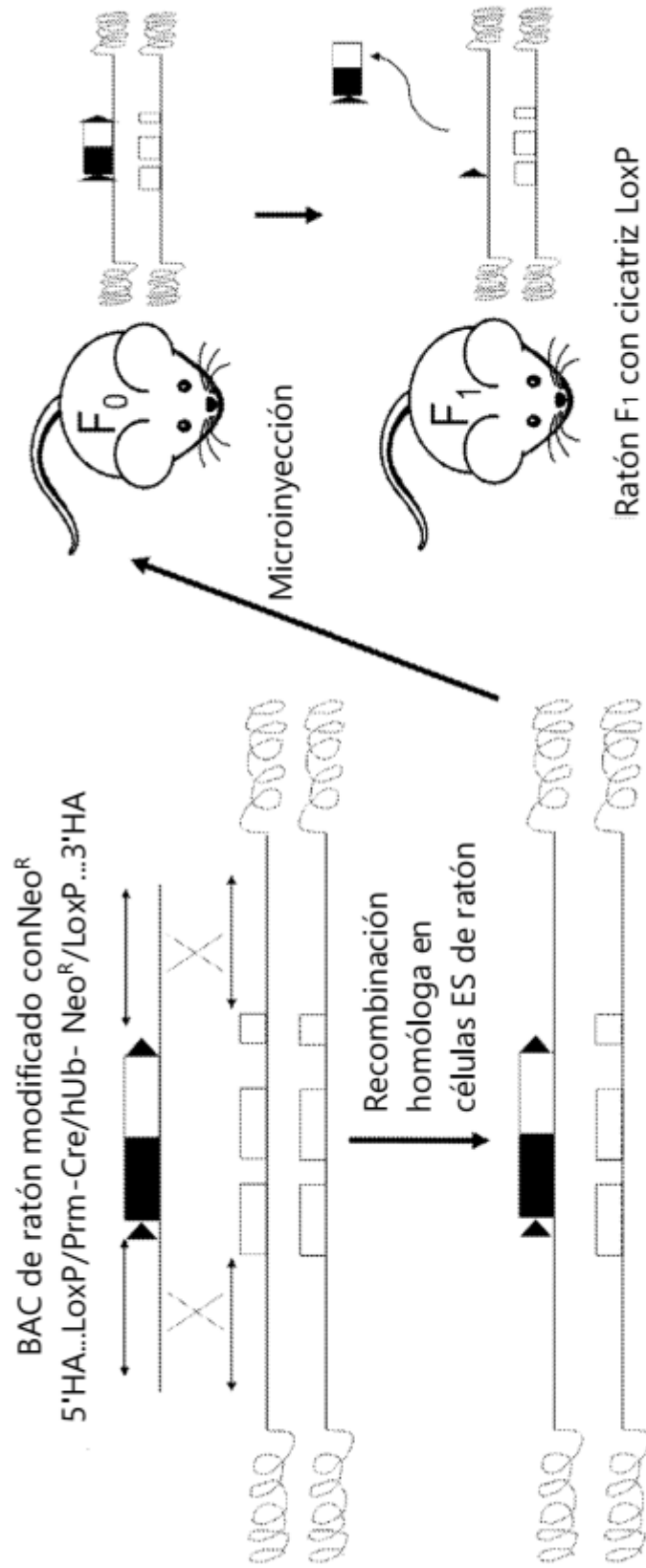


**FIGURA 2**





**FIGURA 4A**



**FIGURA 4B**