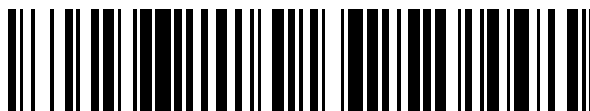


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 947 714**

51 Int. Cl.:

**A01K 67/027** (2006.01)

**C12N 9/22** (2006.01)

**C12N 15/85** (2006.01)

**C12N 15/90** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.12.2015 E 19201384 (5)**

97 Fecha y número de publicación de la concesión europea: **07.06.2023 EP 3653048**

54 Título: **Métodos y composiciones para la modificación genética dirigida a través de múltiples direccionamientos en un solo paso**

30 Prioridad:

**19.12.2014 US 201462094104 P**

**28.05.2015 US 201562167408 P**

**14.08.2015 US 201562205524 P**

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

**17.08.2023**

73 Titular/es:

**REGENERON PHARMACEUTICALS, INC.**

**(100.0%)**

**777 Old Saw Mill River Road**

**Tarrytown, NY 10591, US**

72 Inventor/es:

**VORONINA, VERA;**

**MACDONALD, LYNN;**

**ZAMBROWICZ, BRIAN y**

**MURPHY, ANDREW**

74 Agente/Representante:

**SÁEZ MAESO, Ana**

**ES 2 947 714 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 y composiciones para la modificación genética dirigida a través de múltiples direccionamientos en un solo paso

5

Antecedentes

La recombinación homóloga usando vectores de direccionamiento diseñados para agregar, eliminar o reemplazar una secuencia de ácido nucleico particular en un locus genómico es un enfoque popular para lograr una modificación genómica deseada en animales no humanos.

10

Aunque el estado de la técnica sobre la modificación del genoma a través de la recombinación homóloga ha avanzado considerablemente durante las últimas dos décadas, aún persisten dificultades para lograr una frecuencia de direccionamiento aceptable utilizando vectores de direccionamiento muy grandes, LTVEC, en muchas circunstancias, por ejemplo, cuando una gran parte de un genoma de roedor se reemplaza con un gran fragmento genómico humano, o se dirige a ciertos tipos de células, por ejemplo, fibroblastos u otras células somáticas.

15

Resumen

La presente invención se refiere a las realizaciones caracterizadas en las reivindicaciones. En particular, se proporcionan métodos para modificar un locus genómico diana en una célula de mamífero a través de un sistema de direccionamiento que utiliza dos o más vectores de direccionamiento que son capaces de recombinarse entre sí para formar un segmento de ácido nucleico contiguo único, en los que los vectores de direccionamiento son grandes vectores de direccionamiento (LTVEC) y los LTVEC tienen cada uno al menos un tamaño de 10 kb.

20

25

La invención proporciona métodos de direccionamiento dobles para modificar un locus genómico diana en una célula, que comprende (a) introducir en la célula un agente de nucleasa que produce una ruptura de cadena sencilla o doble dentro del locus genómico diana; (b) introducir en la célula un primer vector de direccionamiento grande (LTVEC) que tiene un tamaño de al menos 10 kb y comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', y un segundo LTVEC que tiene al menos 10 kb de longitud y comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, y el primer brazo de homología 5' del primer LTVEC y el segundo brazo de homología 3' del segundo LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana, en el que el locus genómico diana es modificado por integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula diana que comprenda el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrados en el locus genómico diana.

30

35

40

Opcionalmente, el primer inserto nucleico y el primer brazo de homología 3' y el segundo inserto de ácido nucleico y el segundo brazo de homología 5' son fragmentos superpuestos de un ácido nucleico contiguo, que se reforma mediante la integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico en el locus genómico diana.

45

De acuerdo con los métodos de la invención, la célula es una célula de mamífero. En algunos métodos de este tipo, la célula es una célula humana. En otros de tales métodos, la célula es una célula no humana. En algunos de estos métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal o una célula fibroblástica. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). Opcionalmente, la célula de mamífero es una célula de roedor. Opcionalmente, la célula de roedor es una célula de ratón o una célula de rata.

50

En algunos de los métodos anteriores, el agente de nucleasa es una nucleasa con dedos de cinc (ZFN), una nucleasa efectora similar a un activador de la transcripción (TALEN) o una meganucleasa. En algunos de los métodos anteriores, el agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas regularmente espaciadas (CRISPR) y un ARN guía (ARNg). Opcionalmente, la proteína Cas es Cas9.

55

En algunos métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos, son de una especie que es diferente de la especie de la célula. En algunos métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos, son ácidos nucleicos humanos.

60

En algunos métodos, el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 50 kb a aproximadamente 300 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 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

65

aproximadamente 250 kb a aproximadamente 275 kb, de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 450 kb, o de aproximadamente 450 kb a aproximadamente 500 kb. Opcionalmente, el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 500 kb. Opcionalmente, el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de aproximadamente 300 kb.

En algunos métodos, la célula diana comprende ADN genómico que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico juntos, que tienen un tamaño combinado que oscila entre aproximadamente 5 kb a aproximadamente 500 kb.

En algunos métodos, la primera secuencia superpuesta de la primera LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99.5 %, o 99.9% idéntica a la primera secuencia superpuesta de la segunda LTVEC. De acuerdo con los métodos de la invención, el tamaño de la primera secuencia superpuesta es de al menos 1 kb a aproximadamente 70 kb. En algunos métodos, el tamaño de la primera secuencia superpuesta es de al menos 10 kb o al menos 20 kb.

En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico o ambos en el locus genómico diana da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico diana; (b) una eliminación de una secuencia endógena en el locus genómico diana; o (c) una inserción, una eliminación, una mutación puntual, un intercambio de dominio, un intercambio de exón, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de los mismos. Opcionalmente, la eliminación de la secuencia endógena en el locus genómico diana es 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, o de aproximadamente 700 kb a aproximadamente 800 kb.

En algunos métodos, el uso combinado del primer LTVEC y el segundo LTVEC da como resultado una mayor eficiencia de direccionamiento en comparación con el uso de un solo LTVEC. Opcionalmente, el aumento en la eficiencia de direccionamiento es de al menos 1.5 veces, 2 veces, 2.5 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 11 veces, 12 veces, 13 veces, 14 veces, 15 veces, 16 veces, 17 veces, 18 veces, 19 veces o 20 veces.

En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC o del segundo LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb.

El método de la invención también puede comprender: (a) introducir una célula ES no humana en un embrión huésped no humano, en la que la célula ES no humana se produjo mediante cualquiera de los métodos anteriores. El animal no humano es un ratón o una rata.

La invención también proporciona métodos de direccionamiento triple para modificar un locus genómico diana en una célula, que comprende: (a) introducir en la célula un agente de nucleasa que produce una ruptura de cadena sencilla o doble dentro del locus genómico diana; (b) introducir en la célula un primer vector de direccionamiento grande (LTVEC) que tiene al menos 10 kb de longitud y comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', un segundo LTVEC que tiene al menos 10 kb de longitud y comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un

segundo brazo de homología 3', y un tercer LTVEC que tiene al menos 10 kb de longitud y comprende un tercer inserto de ácido nucleico flanqueado por un tercer brazo de homología 5' y un tercer brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia superpuesta homóloga al tercer brazo de homología 5' del tercer LTVEC, y el primer brazo de homología 5' del primer LTVEC y el tercer brazo de homología 3' del tercer LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana, en el que el locus genómico diana se modifica mediante la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula diana que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrados en el locus genómico diana.

Opcionalmente, el primer inserto nucleico y el primer brazo de homología 3' y el segundo inserto de ácido nucleico y el segundo brazo de homología 5' son fragmentos superpuestos de un ácido nucleico contiguo, y el segundo inserto nucleico y el segundo brazo de homología 3' y el tercer inserto de ácido nucleico y el tercer brazo de homología 5' son fragmentos superpuestos del ácido nucleico contiguo, que se reforma mediante la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico en el locus genómico diana.

En algunos de estos métodos, la célula es una célula humana. En otros métodos de este tipo, la célula es una célula no humana. En algunos de estos métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal o una célula fibroblástica. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). En algunos de estos métodos, la célula es una célula de mamífero. Opcionalmente, la célula de mamífero es una célula de roedor. Opcionalmente, la célula de roedor es una célula de ratón o una célula de rata.

En algunos de estos métodos, el agente nucleasa es una nucleasa con dedos de zinc (ZFN), una nucleasa efectora similar a un activador de la transcripción (TALEN), o una meganucleasa. En algunos de estos métodos, el agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas regularmente interesparadas (CRISPR) y un ARN guía (ARNg). Opcionalmente, la proteína Cas es Cas9.

En algunos de tales métodos, uno o más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico son de una especie que es diferente de la especie de la célula. En algunos de tales métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico son ácidos nucleicos humanos.

En algunos de estos métodos, el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico es de aproximadamente 50 kb a aproximadamente 700 kb, de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 50 kb a aproximadamente 300 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 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, de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 650 kb, o de aproximadamente 650 kb a aproximadamente 700 kb. Opcionalmente, el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 700 kb. Opcionalmente, el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico es de aproximadamente 400 kb.

En algunos de estos métodos, la célula diana comprende ADN genómico que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico juntos, que tienen un tamaño combinado que oscila entre aproximadamente 5 kb y aproximadamente 700 kb.

En algunos de estos métodos, la primera secuencia superpuesta de la primera LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99.5 %, o 99.9 % idéntica a la primera secuencia superpuesta de la segunda LTVEC, y/o la segunda secuencia superpuesta de la segunda LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97%, 98%, 99%, 99.5 %, o 99.9 % idéntica a la segunda secuencia superpuesta de la tercera LTVEC. De acuerdo con los métodos de la invención, el tamaño de la primera secuencia superpuesta es de al menos 1 kb y el tamaño de la segunda secuencia superpuesta es de al menos 1 kb. En algunos de tales métodos, el tamaño de la primera secuencia superpuesta es de aproximadamente 1 kb a aproximadamente 70 kb, y/o el tamaño de la segunda secuencia superpuesta es de aproximadamente 1 kb a aproximadamente 70 kb. En algunos de estos métodos, el tamaño de la primera secuencia superpuesta es de al menos 10 kb o al menos 20 kb, y/o el tamaño de la segunda secuencia superpuesta es de al menos 10 kb o al menos 20 kb.

En algunos de estos métodos, la integración de uno o más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico en el locus genómico diana da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico diana; (b) una eliminación de una secuencia endógena en el locus genómico diana; o (c) una inserción, una eliminación, una mutación puntual, un intercambio de dominio, un intercambio de exón, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de los mismos. Opcionalmente, la eliminación de la secuencia endógena en el locus genómico diana es 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, o de aproximadamente 700 kb a aproximadamente 800 kb.

En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC, el segundo LTVEC o el tercer LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb, la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb, y la suma total de los brazos de homología 5' y 3' del tercer LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb. En algunos de estos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC, el segundo LTVEC o el tercer LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb; la suma total de los brazos de homología 5' y 3' del segundo LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb; y la suma total de los brazos de homología 5' y 3' del tercer LTVEC es 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 120 kb, o de aproximadamente 120 kb a aproximadamente 150 kb.

El método de la invención también puede comprender: (a) introducir una célula ES no humana en un embrión huésped no humano, en el que la célula ES no humana se produjo mediante cualquiera de los métodos anteriores. El animal no humano es un ratón o una rata.

Los métodos y composiciones se proporcionan para modificar un locus genómico diana dentro de una célula de mamífero a través de un sistema de direccionamiento que utiliza dos o más vectores de direccionamiento que son capaces de recombinarse entre sí para formar un segmento de ácido nucleico contiguo único, en el que los vectores de direccionamiento son vectores de direccionamiento grandes (LTVEC), en los que los LTVEC tienen cada uno al menos 10 kb de tamaño.

#### Breve descripción de los dibujos

La Figura 1 proporciona un esquema para un evento de doble direccionamiento genómico en el que se direcciona una célula que tiene una modificación heterocigota del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina. El casete de selección de higromicina es cortado por una nucleasa con dedos de zinc (ZFN) o por un complejo CRISPR/Cas y se direcciona con dos grandes vectores de direccionamiento que comprenden un casete de selección de neomicina y más de 280 kb de segmentos de genes variables de cadena kappa de inmunoglobulina humana. Cada uno de los grandes vectores de direccionamiento comprende una secuencia superpuesta de aproximadamente 20 kb, lo que permite la recombinación homóloga entre los grandes vectores de direccionamiento. El evento de direccionamiento insertó con precisión los segmentos de genes variables de la cadena kappa de inmunoglobulina humana de ambos vectores de direccionamiento en una sola etapa de direccionamiento. Las ubicaciones de las diversas sondas utilizadas para confirmar el evento de direccionamiento se muestran como rectángulos dentro de un círculo. La secuencia de ratón se representa mediante sombreado diagonal ascendente, la secuencia humana se representa sin sombreado, y los sitios de recombinación y casetes de selección se representan mediante sombreado diagonal descendente discontinuo. El esquema no está a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

La Figura 2 proporciona un esquema para un único evento de direccionamiento en el que una célula que tiene una modificación heterocigota del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina se dirige con un vector de direccionamiento grande que comprende un casete de selección de neomicina y 120 kb de segmentos de genes variables de cadena kappa de inmunoglobulina humana. Las ubicaciones de las diversas sondas utilizadas para confirmar el evento de direccionamiento se muestran como rectángulos dentro de un círculo. La secuencia de ratón se representa mediante sombreado diagonal ascendente, la secuencia humana se representa sin sombreado, y los sitios de recombinación y casetes de selección se representan mediante sombreado diagonal descendente discontinuo. El esquema no está a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

La Figura 3 proporciona un esquema para el direccionamiento y la destrucción de un casete de selección de higromicina usando un sistema CRISPR/Cas9 e ilustra la posición dentro del gen de higromicina de los sitios de reconocimiento de CRISPR para los diversos ARN guía (ARNg) que se dirigen a diferentes secuencias en el gen de la higromicina. El esquema no está a escala.

La Figura 4 proporciona un esquema para un evento de triple direccionamiento genómico en el que se dirige una célula que tiene una modificación heterocigota del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina. El casete de selección de higromicina es cortado por una nucleasa con dedos de zinc (ZFN) o por un complejo CRISPR/Cas y se dirige con tres grandes vectores de selección que comprenden un casete de selección de neomicina y aproximadamente 370 kb de segmentos de genes variables de cadena kappa de inmunoglobulina humana. Cada uno de los vectores de direccionamiento grandes comprende una secuencia superpuesta de aproximadamente 20 kb a aproximadamente 60 kb, lo que permite la recombinación homóloga entre los vectores de direccionamiento grandes. El evento de direccionamiento insertó con precisión los segmentos de genes variables de la cadena kappa de inmunoglobulina humana de los tres vectores de direccionamiento en una sola etapa de direccionamiento. Las ubicaciones de las diversas sondas utilizadas para confirmar el evento de direccionamiento se muestran como rectángulos dentro de un círculo. La secuencia de ratón se representa mediante sombreado diagonal ascendente, la secuencia humana se representa sin sombreado, y los sitios de recombinación y casetes de selección se representan mediante sombreado diagonal descendente discontinuo. El esquema no está a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

#### Definiciones

Los términos "proteína", "polipéptido" y "péptido", utilizados en este documento de manera intercambiable, incluyen formas poliméricas de aminoácidos de cualquier longitud, incluidos aminoácidos codificados y no codificados y aminoácidos modificados o derivatizados química o bioquímicamente. Los términos también incluyen polímeros que han sido modificados, tales como polipéptidos que tienen estructuras peptídicas modificadas.

Los términos "ácido nucleico" y "polinucleótido", usados en este documento de manera intercambiable, incluyen formas poliméricas de nucleótidos de cualquier longitud, incluidos ribonucleótidos, desoxirribonucleótidos, o análogos o versiones modificadas de los mismos. Incluyen ADN o ARN monocatenario, bicatenario y multicatenario, ADN genómico, ADNc, híbridos de ADN-ARN y polímeros que comprenden bases de purina, bases de pirimidina u otras bases naturales, químicamente modificadas, bioquímicamente modificadas, no naturales o bases de nucleótidos derivatizadas.

Un locus genómico diana significa una región de un genoma que se va a modificar mediante una modificación dirigida con un vector de direccionamiento. La región se puede definir como la región dentro de los bordes exteriores de los segmentos de ADN genómico correspondientes a los brazos de homología dentro del vector de direccionamiento. Un locus genómico diana puede incluir cualquiera o todos los genes o grupos de genes, uno o más intrones, uno o más exones, una o más secuencias reguladoras y similares.

La "optimización de codones" generalmente incluye un proceso de modificación de una secuencia de ácido nucleico para una expresión mejorada en células huésped particulares mediante la sustitución de al menos un codón de la secuencia nativa con un codón que se usa más frecuentemente o es el más frecuentemente usado en los genes de la célula huésped manteniendo la secuencia de aminoácidos nativa. Por ejemplo, un ácido nucleico que codifica una proteína Cas puede modificarse para sustituir codones que tienen una mayor frecuencia de uso en una célula procarionota o eucariota determinada, incluida 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 natural. Las tablas de uso de codones están fácilmente disponibles, por ejemplo, en la "Codon Usage Database". Estas tablas se pueden adaptar de varias maneras. Véase Nakamura et al. (2000) *Nucleic Acids Research* 28:292. También están disponibles algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en un huésped particular (véase, por ejemplo, Gene Forge).

"Enlace operable" 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 una función que se ejerce sobre al menos uno

de los otros componentes. Por ejemplo, un promotor puede unirse 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.

5 El término "célula pluripotente" o "célula madre pluripotente" incluye una célula no diferenciada que posee la capacidad de convertirse en más de un tipo de célula diferenciada. Tales células pluripotentes pueden ser, por ejemplo, una célula madre embrionaria de mamífero (célula ES) o una célula madre pluripotente inducida de mamífero (célula iPS).

10 El término "célula madre embrionaria" o "célula ES" incluye una célula totipotente o pluripotente derivada de embrión no humano que es capaz de proliferación indiferenciada *in vitro*, y es capaz de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión no humano.

15 El término "célula madre pluripotente inducida" o "célula iPS" incluye una célula madre pluripotente que se puede derivar directamente de una célula adulta diferenciada. Las células iPS humanas se pueden generar introduciendo conjuntos específicos de factores de reprogramación en una célula no pluripotente que pueden incluir, por ejemplo, Oct3/4, factores de transcripción de la familia Sox (p. ej., Sox1, Sox2, Sox3, Sox15), factores de transcripción de la familia Myc (p. ej., c-Myc, 1-Myc, n-Myc), factores de transcripción de la familia tipo Krüppel (KLF) (p. ej., KLF1, KLF2, KLF4, KLF5) y/o factores de transcripción relacionados, tales como NANOG, LIN28, y/o Glis1. Las células iPS humanas también se pueden generar, por ejemplo, mediante el uso de miARN, moléculas pequeñas que imitan las acciones de los factores de transcripción o especificadores de linaje. Las células iPS humanas se caracterizan por su capacidad para diferenciarse en cualquier célula de las tres capas germinales de vertebrados, por ejemplo, el endodermo, el ectodermo o el mesodermo. Las células iPS humanas también se caracterizan por su capacidad de propagarse indefinidamente en condiciones de cultivo *in vitro* adecuadas. Véase, por ejemplo, Takahashi y Yamanaka (Cell (2006) Vol. 126(4), págs. 663-676).

25 El término "línea germinal" en referencia a una secuencia de ácido nucleico de inmunoglobulina incluye una secuencia de ácido nucleico que se puede pasar a la progenie.

30 "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 suelen ser A con T y C con G. En el ARN, suelen ser 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 está unida a una base complementaria mediante el emparejamiento de Watson-Crick. Complementariedad "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 junto con condiciones de hibridación (por ejemplo, concentración de sal y temperatura). Dichas condiciones se pueden predecir usando secuencias y cálculos matemáticos estándar para predecir la T<sub>m</sub> (temperatura de fusión) de las cadenas hibridadas, o mediante la determinación empírica de T<sub>m</sub> usando métodos de rutina. T<sub>m</sub> 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 sencillas). A una temperatura por debajo de la T<sub>m</sub>, se favorece la formación de un complejo de hibridación, mientras que a una temperatura por encima de la T<sub>m</sub>, se favorece la fusión o separación de las cadenas en el complejo de hibridación. La T<sub>m</sub> se puede estimar para un ácido nucleico que tenga un contenido conocido de G+C en una solución acuosa de NaCl 1 M utilizando, por ejemplo,  $T_m = 81.5 + 0.41 (\% \text{ de G+C})$ , aunque otros cálculos de T<sub>m</sub> conocidos tienen en cuenta las características estructurales del ácido nucleico.

50 "Condición de hibridación" incluye el entorno acumulativo en el que una cadena de ácido nucleico se une a una segunda cadena de ácido nucleico mediante interacciones de cadena complementaria y enlaces de hidrógeno para producir un complejo de hibridación. Tales condiciones incluyen los componentes químicos y sus concentraciones (p. ej., sales, agentes quelantes, formamida) de una solución acuosa u orgánica que contiene los ácidos nucleicos y la temperatura de la mezcla. Otros factores, como la duración del tiempo de incubación o las dimensiones de la cámara de reacción, pueden contribuir al medio ambiente. Véase, por ejemplo, Sambrook et al., Molecular Cloning, A Laboratory Manual, 2da ed., páginas 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

60 La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles los desajustes 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 bien conocidas en la técnica. Cuanto mayor sea el grado de complementación entre dos secuencias de nucleótidos, mayor será el valor de la temperatura de fusión (T<sub>m</sub>) 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 desajustes se vuelve importante (véase Sambrook et al., citado anteriormente, 11.7-11.8). Normalmente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico

5 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 se pueden ajustar según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

10 La secuencia de polinucleótido no necesita ser 100% complementaria a la de su ácido nucleico diana para ser hibridable específicamente. Además, un polinucleótido puede hibridar sobre uno o más segmentos de modo que los segmentos intermedios o adyacentes no estén involucrados en el evento de hibridación (p. ej., una estructura de bucle o una estructura de horquilla). Un polinucleótido (p. ej., 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 los que van dirigidos. 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.

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

25 Los métodos y composiciones proporcionados en este documento emplean una variedad de componentes diferentes. Se reconoce a lo largo de la descripción que algunos componentes pueden tener variantes y fragmentos activos. Dichos componentes incluyen, por ejemplo, agentes de nucleasa, proteínas Cas, ARN CRISPR, ARNtracr y ARN guía. La actividad biológica de cada uno de estos componentes se describe en otra parte del presente documento.

30 "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 proteínas, se reconoce que las posiciones de los residuos que no son idénticas a menudo 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 cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservadoras, el porcentaje de identidad de secuencia puede ajustarse hacia arriba para corregir la naturaleza conservadora de la sustitución. Se dice que las secuencias que difieren por tales sustituciones conservadoras tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos. Por lo general, esto implica calificar una sustitución conservadora como un desajuste parcial en lugar de total, aumentando así el porcentaje de identidad de secuencia. Así, 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 puntuación de las sustituciones conservadoras se calcula, p. ej., tal como se implementó en el programa PC/GENE (IntelliGenetics, Mountain View, California).

45 El "porcentaje de identidad de secuencia" incluye el valor determinado comparando dos secuencias alineadas de forma óptima en una ventana de comparación, en la que la parte de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o eliminaciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que aparece la base de ácido nucleico o el residuo de aminoácido idénticos en ambas secuencias para obtener el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicando el resultado por 100 para obtener el porcentaje de identidad de secuencia.

55 A menos que se indique lo contrario, los valores de identidad/similitud de secuencia incluyen el valor obtenido usando GAP Versión 10 usando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando Ponderación de GAP de 50 y Ponderación de Longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos utilizando Ponderación de GAP de 8 y Ponderación de Longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente del mismo. El "programa equivalente" incluye cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genera una alineación que tiene coincidencias idénticas de residuos de nucleótidos o aminoácidos y un porcentaje idéntico de identidad de secuencia en comparación con la alineación correspondiente generada por GAP Versión 10.

65 Una secuencia "homóloga" incluye una secuencia de ácido nucleico que es idéntica o sustancialmente similar a una secuencia de referencia conocida, tal que es al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 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. Una secuencia "ortóloga" incluye una secuencia de ácido nucleico de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie.

5 El término "*in vitro*" incluye entornos artificiales y procesos o reacciones que ocurren dentro de un entorno artificial (por ejemplo, un tubo de ensayo). El término "*in vivo*" incluye entornos naturales (por ejemplo, una célula, un organismo o un cuerpo) y procesos o reacciones que ocurren dentro de un entorno natural. El término "*ex vivo*" incluye células que han sido extraídas del cuerpo de un individuo y procesos o reacciones que ocurren dentro de dichas células.

10 Las composiciones o métodos que "comprenden" o "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.

15 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 a partir del contexto, el término "aproximadamente" abarca valores dentro de un margen estándar de error de medición (por ejemplo, SEM) de un valor establecido.

20 Las formas singulares de los artículos "un", "uno, una" y "el, la" incluyen referencias en plural a menos que el contexto dicte claramente lo contrario. Por ejemplo, el término "una proteína Cas" o "al menos una proteína Cas" puede incluir una pluralidad de proteínas Cas, incluidas sus mezclas.

25 Descripción detallada

I. Modificación de loci genómicos utilizando múltiples vectores de direccionamiento

30 La presente invención se relaciona con métodos para modificar un locus genómico diana dentro de una célula de mamífero como se caracteriza en las reivindicaciones. Dichos métodos emplean múltiples vectores de direccionamiento grandes (LTVEC) que son capaces de recombinarse entre sí para formar un único segmento de ácido nucleico contiguo. Dichos métodos pueden utilizar 1, 2, 3, 4, 5, 6 o más LTVEC en una sola etapa de direccionamiento. También se proporcionan métodos y composiciones para mejorar la recombinación homóloga en un locus genómico diana en una célula. Dichos métodos emplean dos o más ácidos nucleicos que comprenden una o más secuencias superpuestas. Cualquiera de los métodos divulgados en el presente documento puede tener lugar *in vitro*, *ex vivo* o *in vivo*.

A. Direccionamiento doble

40 Se proporcionan métodos para modificar un locus genómico diana dentro de una célula de mamífero a través de un método de direccionamiento doble como se caracteriza en las reivindicaciones. Los métodos emplean dos vectores de direccionamiento grandes (LTVEC) (es decir, un primer LTVEC y un segundo LTVEC) que son capaces de recombinarse entre sí para formar un único segmento de ácido nucleico contiguo. El primer LTVEC comprende un primer inserto de ácido nucleico y el segundo LTVEC comprende un segundo inserto de ácido nucleico. Los insertos de ácido nucleico están flanqueados por brazos de homología 5' y 3'. El primer inserto de ácido nucleico y su brazo de homología 3' y el segundo inserto de ácido nucleico y su brazo de homología 5' pueden ser fragmentos superpuestos del mismo ácido nucleico contiguo. El brazo de homología 3' del primer LTVEC y el brazo de homología 5' del segundo LTVEC se superponen (es decir, son complementarios entre sí) y las inserciones primera y segunda flanquean los brazos de homología superpuestos. Dichos métodos involucran tres eventos de recombinación que pueden ocurrir en cualquier orden: (1) recombinación entre el brazo de homología 3' del primer LTVEC y el brazo de homología 5' del segundo LTVEC; (2) recombinación entre el brazo de homología 5' del primer LTVEC y el segmento correspondiente en el locus diana; y (3) recombinación entre el brazo de homología 3' del segundo LTVEC y el segmento correspondiente en el locus diana. Esta recombinación de tres vías reconstruye el ácido nucleico contiguo en el locus diana con la secuencia superpuesta de los brazos de homología posicionados entre el primer y el segundo insertos de ácido nucleico.

55 Cada uno de los LTVEC también comprende un brazo de homología 5' o 3' que es homólogo a una región de ADN dentro o cerca de un locus genómico diana que permite la recombinación e integración del único segmento de ácido nucleico contiguo. Por lo tanto, por medio de un evento de recombinación de tres vías, se puede realizar una gran modificación de ácido nucleico (es decir, eliminación, inserción y/o reemplazo) en un locus diana en una sola etapa de direccionamiento.

60 Los tres eventos de recombinación pueden ocurrir en cualquier orden. En una realización, el evento de recombinación entre las secuencias superpuestas de los dos LTVEC ocurre antes de la recombinación homóloga con el locus diana. En otra realización, la recombinación con el locus diana ocurre antes de la recombinación entre los dos LTVEC. En otra realización más, los tres eventos de recombinación pueden ocurrir simultáneamente.

En una realización, se proporciona un método para modificar un locus genómico diana en una célula. Dicho método comprende la introducción de un primer vector de direccionamiento grande (LTVEC) que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', y un segundo LTVEC que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, y el primer brazo de homología 5' del primer LTVEC y el segundo brazo de homología 3' del segundo LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana; en el que el locus genómico diana se modifica mediante la integración de los insertos de ácido nucleico primero y segundo entre los segmentos genómicos correspondientes. El método comprende además seleccionar una célula diana que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrados en el locus genómico diana.

#### B. Direccionamiento triple

También se proporcionan métodos para modificar un locus genómico diana dentro de una célula mediante métodos de direccionamiento triple. Los métodos emplean tres vectores de direccionamiento grandes (LTVEC) (es decir, un primer LTVEC, un segundo LTVEC y un tercer LTVEC) que son capaces de recombinarse entre sí para formar un único segmento de ácido nucleico contiguo. El primer LTVEC comprende un primer inserto de ácido nucleico, el segundo LTVEC comprende un segundo inserto de ácido nucleico y el tercer LTVEC comprende un tercer inserto de ácido nucleico. Los insertos de ácido nucleico están flanqueados por brazos de homología 5' y 3'. El primer inserto de ácido nucleico y su brazo de homología 3' y el segundo inserto de ácido nucleico y su brazo de homología 5' pueden ser fragmentos superpuestos del mismo ácido nucleico contiguo. El segundo inserto de ácido nucleico y su brazo de homología 3' y el tercer inserto de ácido nucleico y su brazo de homología 5' pueden ser fragmentos superpuestos del mismo ácido nucleico contiguo. El brazo de homología 3' del primer LTVEC y el brazo de homología 5' del segundo LTVEC se superponen (es decir, son complementarios entre sí) y los insertos primero y segundo flanquean los brazos de homología superpuestos. El brazo de homología 3' del segundo LTVEC y el brazo de homología 5' del tercer LTVEC se superponen (es decir, son complementarios entre sí) y los insertos segundo y tercero flanquean los brazos de homología superpuestos.

Dichos métodos implican cuatro eventos de recombinación que pueden ocurrir en cualquier orden: (1) recombinación entre el brazo de homología 3' del primer LTVEC y el brazo de homología 5' del segundo LTVEC; (2) recombinación entre el brazo de homología 3' del segundo LTVEC y el brazo de homología 5' del tercer LTVEC; (3) recombinación entre el brazo de homología 5' del primer LTVEC y el segmento correspondiente en el locus diana; y (4) recombinación entre el brazo de homología 3' del tercer LTVEC y el segmento correspondiente en el locus diana. Esta recombinación de cuatro vías reconstruye el ácido nucleico contiguo en el locus diana con la secuencia superpuesta de los brazos de homología colocados entre el primer y el segundo insertos de ácido nucleico y entre el segundo y el tercer insertos de ácido nucleico.

El primer y tercer LTVEC también comprenden un brazo de homología 5' o 3' que es homólogo a una región de ADN dentro o cerca de un locus genómico diana, que permite la recombinación e integración del único segmento de ácido nucleico contiguo. Por lo tanto, por medio de un evento de recombinación de cuatro vías, se puede realizar una gran modificación de ácido nucleico (es decir, eliminación, inserción y/o reemplazo) en un locus diana en una sola etapa de direccionamiento.

Los cuatro eventos de recombinación pueden ocurrir en cualquier orden. En una realización, el evento de recombinación entre las secuencias superpuestas de las tres LTVEC ocurre antes de la recombinación homóloga con el locus diana. En otra realización, la recombinación con el locus diana ocurre antes de la recombinación entre los tres LTVEC. En otra realización más, los cuatro eventos de recombinación pueden ocurrir simultáneamente.

En una realización, se proporciona un método para modificar un locus genómico diana en una célula. Dicho método comprende la introducción de un primer vector de direccionamiento grande (LTVEC) que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', un segundo LTVEC que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y un tercer LTVEC que comprende un tercer inserto de ácido nucleico flanqueado por un tercer brazo de homología 5' y un tercer brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una secuencia superpuesta homóloga al tercer brazo de homología 5' del tercer LTVEC, y el primer brazo de homología 5' del primer LTVEC y el tercer brazo de homología 3' del tercer LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana; en el que el locus genómico diana se modifica mediante la integración de los insertos de ácido nucleico primero, segundo y tercero entre los segmentos genómicos correspondientes. El método comprende además seleccionar una célula diana que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrados en el locus genómico diana.

#### C. Direccionamiento con múltiples LTVEC

Los métodos de direccionamiento proporcionados en el presente documento para crear una modificación genética en una sola etapa de direccionamiento proporcionan nuevas posibilidades y eficiencias mejoradas para modificaciones dirigidas de genes más allá de las que se logran con un solo método de direccionamiento LTVEC. El direccionamiento con dos, tres o más LTVEC que son capaces de recombinarse entre sí permite la modificación de un segmento más grande de ADN. Los eventos de recombinación pueden ocurrir en cualquier orden. Por ejemplo, el evento de recombinación entre las secuencias superpuestas de las LTVEC puede ocurrir antes de la recombinación homóloga con el locus diana. Alternativamente, la recombinación con el locus diana puede ocurrir antes de que la recombinación entre los LTVEC o los eventos de recombinación puedan ocurrir simultáneamente.

Los métodos de direccionamiento descritos en el presente documento proporcionan varias ventajas sobre los métodos de direccionamiento de LTVEC únicos existentes, que incluyen una mayor eficiencia de direccionamiento, un aumento en el tamaño alcanzable de la modificación genética y una reducción en el número de etapas de direccionamiento necesarios para obtener grandes modificaciones genómicas lo que ahorra tiempo y mantiene la pluripotencia de las células madre embrionarias modificadas. Esto es de particular importancia para las grandes modificaciones genómicas, ya que los métodos permiten la modificación del locus genómico con una combinación de insertos de ácido nucleico de dos, tres o más LTVEC en una sola etapa. Por lo tanto, tales modificaciones pueden permitir eliminaciones, reemplazos e inserciones muy grandes (p. ej., > 50 kb) dentro del locus genómico diana.

Por ejemplo, el tiempo requerido para usar tres LTVEC de manera secuencial para modificar un locus genómico diana y detectar y confirmar la modificación dirigida es de aproximadamente nueve meses, mientras que la misma modificación se puede realizar y confirmar con tres LTVEC simultáneamente en sólo unos cuatro meses.

Las modificaciones secuenciales también crean un mayor riesgo de pérdida de pluripotencia y potencial de transmisión de línea germinal cuando se modifican células pluripotentes tales como células madre embrionarias. A medida que aumenta el número de pases en cultivo y aumenta el número de electroporaciones, se acumulan anomalías cromosómicas y cariotípicas y pueden causar una pérdida de competencia de la línea germinal. Véase, por ejemplo, Buehr et al. (2008) Cell 135: 1287-1298; Li et al. (2008) Cell 135(7): 1299-1310; y Liu et al. (1997) Dev. Dyn. 209: 85-91. El uso de múltiples LTVEC simultáneamente en lugar de secuencialmente reduce el número de pases y el número de electroporaciones y, por lo tanto, aumenta la capacidad de realizar manipulaciones genéticas en células pluripotentes, como las células madre embrionarias, al tiempo que conserva su competencia de línea germinal.

En realizaciones particulares, la modificación genética comprende una modificación de uno o más ácidos nucleicos endógenos, una sustitución de uno o más ácidos nucleicos endógenos, un reemplazo de un ácido nucleico endógeno por un ácido nucleico heterólogo, una eliminación o una inserción. En ejemplos específicos, la modificación genética se introduce mediante la introducción de al menos dos vectores de direccionamiento grandes (LTVEC) en una célula. En otro ejemplo, la modificación genética se introduce mediante la introducción de al menos tres vectores de direccionamiento grandes (LTVEC) en una célula. En tales ejemplos, los LTVEC pueden comprender ADN que se va a insertar en el locus genómico diana de la célula.

En algunas realizaciones, los métodos para modificar un locus genómico diana comprenden introducir una modificación genética en células de mamífero. Asimismo, la invención proporciona células de mamífero que comprenden una modificación genética.

Se pueden usar varios métodos para realizar modificaciones genéticas dirigidas en las células. Por ejemplo, como se describió anteriormente, la modificación genética dirigida emplea un sistema que generará una modificación genética dirigida a través de un evento de recombinación homóloga. En otros casos, una célula se puede modificar usando agentes de nucleasa que generan una rotura de cadena sencilla o doble en un locus genómico diana. La ruptura de una cadena sencilla o doble luego se repara mediante la vía de unión de extremos no homólogos (NHEJ). Los ejemplos de métodos para generar tales modificaciones genéticas dirigidas se discuten en detalle en otra parte del presente documento, incluido, por ejemplo, el uso de vectores de direccionamiento grandes. Véase también Wang et al. (2013) Cell 153 :910-918, Mandalos et al. (2012) PLOS ONE 7:e45768: 1-9, y Wang et al. (2013) Nat Biotechnol. 31 :530-532.

La modificación de genes dirigidos mediante recombinación homóloga entre un vector de direccionamiento y un locus diana puede ser muy ineficaz, especialmente en tipos de células distintas de las células madre embrionarias de roedores. El uso de un vector de direccionamiento en combinación con una rotura de ADN de doble cadena dirigida por nucleasas en el locus diana puede mejorar en gran medida la eficiencia de direccionamiento para modificaciones, tales como eliminaciones o inserciones. De manera similar, el uso de un vector de direccionamiento en combinación con una rotura de ADN monocatenario dirigida por nucleasa en el locus diana puede mejorar en gran medida la eficiencia de direccionamiento para las modificaciones.

De acuerdo con las realizaciones caracterizadas en las reivindicaciones, los LTVEC se pueden emplear en combinación con agentes de nucleasa que producen una rotura de cadena sencilla o doble dentro de un locus genómico diana. Tal método comprende además introducir un agente de nucleasa en una célula. En una realización, el agente de nucleasa es una nucleasa con dedos de zinc (ZFN). En otra realización, el agente de nucleasa es un compuesto de repeticiones palindrómicas cortas agrupadas regularmente intercaladas. (CRISPR)/sistema asociado a CRISPR (Cas).

En una realización, se proporciona un método de direccionamiento doble para modificar un locus genómico diana en una célula, comprendiendo el método: (a) introducir en una célula un agente de nucleasa que hace que una cadena sencilla o doble se rompa dentro de un locus genómico diana; (b) introducir un primer vector de direccionamiento grande (LTVEC) que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', y un segundo LTVEC que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC y el primer brazo de homología 5' del primer LTVEC y el segundo brazo de homología 3' del segundo LTVEC son homólogos a los correspondientes segmentos genómicos dentro del locus genómico diana; en el que el locus genómico diana se modifica mediante la integración del primer y segundo insertos de ácido nucleico, entre los segmentos genómicos correspondientes; y (c) seleccionar una célula diana que comprende el primer inserto de ácido nucleico, y el segundo inserto de ácido nucleico integrados en el locus genómico diana. En dichos métodos, el primer inserto nucleico y el primer brazo de homología 3' y el segundo inserto de ácido nucleico y el segundo brazo de homología 5' son fragmentos superpuestos de un ácido nucleico contiguo, que se reforma mediante la integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico en el locus genómico diana.

En una realización, se proporciona un método de direccionamiento triple para modificar un locus genómico diana en una célula, comprendiendo el método: (a) introducir en una célula un agente de nucleasa que hace que una cadena sencilla o doble se rompa dentro de un locus genómico diana; (b) introducir un primer vector de direccionamiento grande (LTVEC) que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', un segundo LTVEC que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y un tercer LTVEC que comprende un tercer inserto de ácido nucleico flanqueado por un tercer brazo de homología 5' y un tercer brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una secuencia superpuesta homóloga al tercer brazo de homología 5' del tercer LTVEC, y el primer brazo de homología 5' del primer LTVEC y el tercer brazo de homología 3' del tercer LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana; en el que el locus genómico diana se modifica mediante la integración de los insertos de ácido nucleico primero, segundo y tercero entre los segmentos genómicos correspondientes; y (c) seleccionar una célula diana que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrados en el locus genómico diana. En dichos métodos de direccionamiento triple, el primer inserto nucleico y el primer brazo de homología 3' y el segundo inserto de ácido nucleico y el segundo brazo de homología 5' son fragmentos superpuestos de un ácido nucleico contiguo, y el segundo inserto nucleico y el segundo brazo de homología 3' y el tercer inserto de ácido nucleico y el tercer brazo de homología 5' son fragmentos superpuestos de un ácido nucleico contiguo, que se reforma mediante la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico en el locus genómico diana.

En algunos casos, los dos, tres o más LTVEC se pueden introducir simultáneamente. Alternativamente, los dos, tres o más LTVEC pueden introducirse secuencialmente o pueden introducirse en diferentes momentos.

Los diversos componentes del sistema de direccionamiento pueden incluir, por ejemplo, vectores de direccionamiento, agentes de nucleasa, un locus genómico diana, insertos de ácido nucleico, polinucleótidos de interés y/u otros componentes, cada uno de los cuales se describe en detalle en otra parte del presente documento.

#### D. Direccionamiento con múltiples ácidos nucleicos superpuestos

Los métodos de direccionamiento proporcionados en el presente documento para crear una modificación genética en una sola etapa de direccionamiento brindan nuevas posibilidades y eficiencias mejoradas para modificaciones de genes dirigidos más allá de las logradas con un solo ácido nucleico.

En el presente documento se proporcionan métodos para modificar un locus genómico diana o mejorar la recombinación homóloga en un locus genómico diana en una célula, que comprende introducir en la célula un primer y segundo ácidos nucleicos, en los que el primero y el segundo ácidos nucleicos comprenden una secuencia superpuesta. Los ácidos nucleicos primero y segundo pueden ser, por ejemplo, ácidos nucleicos lineales. Dichos métodos también pueden comprender introducir en la célula tres o más ácidos nucleicos que sean capaces de recombinarse entre sí. Por ejemplo, los ácidos nucleicos primero y segundo pueden tener una primera secuencia superpuesta, y los ácidos nucleicos segundo y tercero pueden tener una segunda secuencia superpuesta. En los métodos de la invención, se modifica el locus genómico diana, o se potencia la recombinación homóloga en el locus genómico diana, con la ayuda de una nucleasa que hace que una cadena sencilla o doble se rompa en o cerca del locus genómico diana, tal como una nucleasa con dedos de zinc, un TALEN, una meganucleasa o Cas9 y un ARN guía.

El método puede mejorar la recombinación homóloga del primer ácido nucleico en el locus genómico diana, puede mejorar la recombinación homóloga del segundo ácido nucleico en el locus genómico diana, o puede mejorar la

recombinación homóloga del primer y segundo ácido nucleico en el locus genómico diana. Como ejemplo, la recombinación homóloga del primer ácido nucleico en el locus genómico diana puede mejorarse en comparación con los métodos en los que el primer ácido nucleico se introduce sin el segundo ácido nucleico. Asimismo, la recombinación homóloga del segundo ácido nucleico en el locus genómico diana puede mejorarse en comparación con los métodos en los que el segundo ácido nucleico se introduce sin el primer ácido nucleico. La mejora de la recombinación homóloga puede ser, por ejemplo, al menos 1.5 veces, 2 veces, 2.5 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 veces o 20 veces. Por ejemplo, el cambio de veces en la mejora con una nucleasa puede ser de 0.5 veces, 0.6 veces, 0.7 veces, 0.8 veces, 0.9 veces, 1.0 veces, 1.1 veces, 1.2 veces, 1.3 veces, 1.4 veces, 1.5 veces, 1.6 veces, 1.7 veces, 1.8 veces, 1.9 veces, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces o 10 veces en comparación con la mejora sin una nucleasa.

La secuencia superpuesta del primer ácido nucleico puede ser homóloga a la secuencia superpuesta del segundo ácido nucleico. Por ejemplo, la secuencia superpuesta del primer ácido nucleico puede ser al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99.5 % o 99.9 % idéntica a la secuencia superpuesta del segundo ácido nucleico.

Alternativamente, la secuencia superpuesta del primer ácido nucleico puede ser 100% idéntica a la secuencia superpuesta del segundo ácido nucleico.

La secuencia superpuesta es de al menos 1 kb y puede ser, por ejemplo, de aproximadamente 1 kb a aproximadamente 70 kb o más. Por ejemplo, la secuencia superpuesta puede ser de aproximadamente 1 kb a aproximadamente 5 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 15 kb, de aproximadamente 15 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 25 kb, de aproximadamente 25 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 35 kb, de aproximadamente 35 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 45 kb, de aproximadamente 45 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 55 kb, de aproximadamente 55 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 65 kb, de aproximadamente 65 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 220 kb, de aproximadamente 220 kb a aproximadamente 240 kb, de aproximadamente 240 kb a aproximadamente 260 kb, de aproximadamente 260 kb a aproximadamente 280 kb, o de aproximadamente 280 kb a aproximadamente 300 kb. Como ejemplo, la secuencia superpuesta puede ser de aproximadamente 20 kb a aproximadamente 60 kb. Alternativamente, la secuencia superpuesta puede ser de al menos 1 kb, al menos 5 kb, al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 25 kb, al menos 30 kb, al menos 35 kb, al menos 40 kb, al menos 45 kb, al menos 50 kb, al menos 55 kb, al menos 60 kb, al menos 65 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 120 kb, al menos 140 kb, al menos 160 kb, al menos 180 kb, al menos 200 kb, al menos 220 kb, al menos 240 kb, al menos 260 kb, al menos 280 kb, o al menos 300 kb.

La secuencia superpuesta se puede ubicar en cualquier lugar dentro del primer y segundo ácidos nucleicos. Por ejemplo, la secuencia superpuesta se puede ubicar en el extremo 3' del primer ácido nucleico y el extremo 5' del segundo ácido nucleico. Alternativamente, la secuencia superpuesta se puede ubicar en el extremo 5' del primer ácido nucleico y en el extremo 3' del segundo ácido nucleico.

El primer ácido nucleico es un vector de direccionamiento que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', y el segundo ácido nucleico es un segundo vector de direccionamiento que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'.

El primer vector de direccionamiento un vector de direccionamiento grande (LTVEC) que es al menos de 10 kb. Asimismo, el segundo vector de direccionamiento es un LTVEC que es al menos de 10 kb. Los ejemplos de tamaños de LTVEC se divulgan en otra parte del presente documento. Por ejemplo, el primer y/o segundo LTVEC puede ser 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 120 kb, de aproximadamente 120 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 650 kb, de aproximadamente 650 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 750 kb, de aproximadamente 750 kb a aproximadamente 800 kb.

En algunos métodos, el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico son fragmentos superpuestos de un ácido nucleico contiguo. En algunos métodos, el primer y/o segundo insertos de ácido nucleico

pueden ser de una especie diferente a la especie de la célula. Por ejemplo, el primer y/o segundo insertos de ácido nucleico pueden ser ácidos nucleicos humanos.

5 Los métodos pueden dar como resultado la integración del primer y/o segundo insertos de ácido nucleico en el locus genómico diana. La integración puede resultar en la adición de una secuencia en el locus genómico diana, la eliminación de una secuencia en el locus genómico diana o el reemplazo de una secuencia en el locus genómico diana. Por ejemplo, la integración puede resultar en la adición de una secuencia exógena en el locus genómico diana, la eliminación de una secuencia endógena en el locus genómico diana o el reemplazo de una secuencia endógena con una secuencia exógena en el locus genómico diana. El primer inserto de ácido nucleico, el segundo inserto de  
10 ácido nucleico o la combinación del primer y segundo insertos de ácido nucleico que se insertan en el locus genómico diana pueden ser, por ejemplo, de aproximadamente 5 kb a aproximadamente 500 kb. Otro ejemplo de inserto de ácido nucleico y de tamaños de inserción se divulgan en otra parte del presente documento. La eliminación en el locus genómico diana puede ser, por ejemplo, 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, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, o de aproximadamente 700 kb a aproximadamente 800 kb. Otros ejemplos de tamaños de supresión se divulgan en otra parte del presente documento.

La célula diana puede ser cualquiera de los tipos de células proporcionados en el presente documento, y el locus genómico diana puede ser cualquier ADN dentro de la célula. Por ejemplo, el locus genómico diana puede estar en el genoma de la célula o puede estar en el ADN extracromosómico dentro de la célula.

25 II. Insertos de ácido nucleico y vectores de direccionamiento

#### A. Inserto de ácido nucleico

30 Se pueden emplear uno o más insertos de ácido nucleico en los métodos divulgados en el presente documento, y se pueden introducir en una célula a través de vectores de direccionamiento separados o en el mismo vector de direccionamiento. Los insertos de ácido nucleico incluyen segmentos de ADN que se integrarán en los loci diana genómicos. La integración de un inserto de ácido nucleico en un locus diana puede resultar en la adición de una secuencia de ácido nucleico de interés al locus diana, la eliminación de una secuencia de ácido nucleico de interés en el locus diana y/o el reemplazo de una secuencia de ácido nucleico de interés en el locus diana.  
35

Los métodos proporcionan la modificación de un locus genómico con insertos de ácido nucleico que tienen un tamaño mayor que el que se puede lograr utilizando técnicas convencionales de direccionamiento único (es decir, un solo LTVEC). En dichos métodos, los insertos de ácido nucleico se incluyen en dos, tres o más LTVEC. Los LTVEC están diseñados de tal manera que son capaces de recombinarse entre sí para formar un solo segmento grande de ADN que comprende los insertos de ácido nucleico combinados de los dos, tres o más LTVEC.  
40

En dichos métodos, los insertos de ácido nucleico están flanqueados por brazos de homología 5' y 3'. El brazo de homología 3' que flanquea el primer inserto de ácido nucleico y el brazo de homología 5' que flanquea el segundo inserto de ácido nucleico son fragmentos superpuestos del mismo ácido nucleico contiguo que luego se reforma por recombinación entre los fragmentos superpuestos de los brazos de homología. En dichos métodos, la recombinación entre los dos LTVEC da como resultado un inserto de ácido nucleico contiguo con la secuencia superpuesta de los brazos de homología posicionados entre el primer y el segundo inserto de ácido nucleico. Los métodos de direccionamiento triple implican una recombinación adicional entre el segundo LTVEC y el tercer LTVEC en el que el brazo de homología 3' que flanquea el segundo inserto de ácido nucleico y el brazo de homología 5' que flanquea el tercer inserto de ácido nucleico son fragmentos superpuestos del mismo ácido nucleico contiguo que luego se reforma por recombinación entre los fragmentos superpuestos de los brazos de homología. En dichos métodos de direccionamiento triple, la recombinación entre los tres LTVEC da como resultado un inserto de ácido nucleico contiguo con la secuencia superpuesta de los brazos de homología posicionados entre el primer, el segundo y el tercer inserto de ácido nucleico. En una realización, la secuencia superpuesta de los brazos de homología comprende una porción del inserto de ácido nucleico.  
45  
50  
55

Como tal, estos métodos permiten una modificación de un locus genómico con insertos de ácido nucleico de dos, tres o más LTVEC en una sola etapa de direccionamiento, aumentando así de manera efectiva el tamaño total del inserto de ácido nucleico mientras que al mismo tiempo reduce el número de etapas de direccionamiento.  
60

El inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor, un potenciador o cualquier combinación de los mismos. Además, el tamaño del inserto de ácido nucleico (es decir, los insertos de ácido nucleico combinados de dos, tres o más LTVEC) o el ácido nucleico correspondiente en el locus diana que se reemplaza puede tener cualquier longitud deseada, incluso, por ejemplo, entre 10-100 nucleótidos de longitud, 100-  
65

500 nucleótidos de longitud, 500 nucleótidos - 1 kb de longitud, 1 kb a 1.5 kb de longitud, 1.5 kb a 2 kb de longitud, 2 kb a 2.5 kb de longitud, 2.5 kb a 3 kb de longitud, 3 kb a 5 kb de longitud, 5 kb a 8 kb de longitud, 8 kb a 10 kb de longitud o más. En otros casos, la longitud puede ser de aproximadamente 50 kb a aproximadamente 700 kb, de aproximadamente 50 kb a aproximadamente 500 kb, de aproximadamente 50 kb a aproximadamente 300 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, de aproximadamente 275 kb a aproximadamente 300 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 650 kb, de aproximadamente 650 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 800 kb, de aproximadamente 800 kb a 1 Mb, de aproximadamente 1 Mb a aproximadamente 1.5 Mb, de aproximadamente 1.5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2.5 Mb, de aproximadamente 2.5 Mb a aproximadamente 2.8 Mb, o de aproximadamente 2.8 Mb a aproximadamente 3 Mb. Alternativamente, las inserciones de ácido nucleico combinadas de los dos, tres o más LTVEC o el ácido nucleico correspondiente en el locus diana que se reemplaza pueden ser de aproximadamente 3 Mb a aproximadamente 4 Mb, de aproximadamente 4 Mb a aproximadamente 5 Mb, de aproximadamente 5 Mb a aproximadamente 6 Mb, de aproximadamente 6 Mb a aproximadamente 7 Mb, de aproximadamente 7 Mb a aproximadamente 8 Mb, de aproximadamente 8 Mb a aproximadamente 9 Mb, o de aproximadamente 9 Mb a aproximadamente 10 Mb. En aún otros casos, la longitud puede ser de al menos 100, 200, 300, 400, 500, 600, 700, 800 o 900 nucleótidos o de al menos 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb o más. Por ejemplo, las inserciones de ácido nucleico combinadas de los dos, tres o más LTVEC o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser de al menos 20 kb, al menos 40 kb, al menos 60 kb, al menos 80 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, al menos 500 kb, al menos 550 kb, al menos 600 kb, al menos 650 kb, al menos 700 kb, al menos 750 kb, al menos 800 kb, al menos 850 kb, al menos 900 kb, al menos 950 kb, al menos 1 Mb, al menos 1.5 Mb, al menos 2 Mb, al menos 2.5 Mb, al menos 3 Mb, al menos 4 Mb, al menos 5 Mb, al menos 6 Mb, al menos 7 Mb, al menos 8 Mb, al menos 9 Mb, al menos 10 Mb. En una realización, el tamaño del inserto de ácido nucleico es de aproximadamente 5 kb a aproximadamente 700 kb. En una realización, el tamaño del inserto de ácido nucleico es de aproximadamente 5 kb a aproximadamente 500 kb. En otra realización, el tamaño del inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 700 kb. En otra realización, el tamaño del inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 500 kb. En una realización específica, el inserto de ácido nucleico es de aproximadamente 140 kb. En otra realización específica, el inserto de ácido nucleico es de aproximadamente 370 kb. En otra realización específica, el inserto de ácido nucleico es de aproximadamente 300 kb. En otra realización específica, el inserto de ácido nucleico es de aproximadamente 400 kb.

En algunos vectores de direccionamiento individuales (es decir, antes de la recombinación con otro vector de direccionamiento), el inserto de ácido nucleico puede tener entre 10 y 100 nucleótidos de longitud, 100 y 500 nucleótidos de longitud, 500 nucleótidos - 1 kb de longitud, 1 kb a 1.5 kb de longitud, 1.5 kb a 2 kb de longitud, 2 kb a 2.5 kb de longitud, 2.5 kb a 3 kb de longitud o 3 kb a 5 kb de longitud. En otros casos, la longitud puede ser de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 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 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb. Alternativamente, el inserto de ácido nucleico puede ser 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. Alternativamente, el inserto de ácido nucleico puede ser de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de

aproximadamente 600 kb a aproximadamente 650 kb, de aproximadamente 650 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 750 kb, o de aproximadamente 750 kb a aproximadamente 800 kb.

5 En algunos casos, el reemplazo del ácido nucleico en el locus diana da como resultado la eliminación de una secuencia de ácido nucleico que varía de aproximadamente 1 kb a aproximadamente 200 kb, de aproximadamente 2 kb a aproximadamente 20 kb, o de aproximadamente 0.5 kb a aproximadamente 3 Mb. En algunos casos, la extensión de la eliminación es mayor que la longitud total del brazo de homología 5' y el brazo de homología 3'.

10 En algunos casos, la extensión de la eliminación de la secuencia de ácido nucleico varía 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 de 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, de aproximadamente 150 kb a aproximadamente 200 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 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, de aproximadamente 190 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 800 kb, de aproximadamente 800 kb a 1 Mb, de aproximadamente 1 Mb a aproximadamente 1.5 Mb, de aproximadamente 1.5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2.5 Mb, de aproximadamente 2.5 Mb a aproximadamente 2.8 Mb, de aproximadamente 2.8 Mb a aproximadamente 3 Mb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 1 Mb, de aproximadamente 1 Mb a aproximadamente 1.5 Mb, de aproximadamente 1.5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2.5 Mb, o de aproximadamente 2.5 Mb a aproximadamente 3 Mb. Alternativamente, la eliminación puede ser de aproximadamente 3 Mb a aproximadamente 4 Mb, de aproximadamente 4 Mb a aproximadamente 5 Mb, de aproximadamente 5 Mb a aproximadamente 10 Mb, de aproximadamente 10 Mb a aproximadamente 20 Mb, de aproximadamente 20 Mb a aproximadamente 30 Mb, de aproximadamente 30 Mb a aproximadamente 40 Mb, de aproximadamente 40 Mb a aproximadamente 50 Mb, de aproximadamente 50 Mb a aproximadamente 60 Mb, de aproximadamente 60 Mb a aproximadamente 70 Mb, de aproximadamente 70 Mb a aproximadamente 80 Mb, de aproximadamente 80 Mb a aproximadamente 90 Mb, o de aproximadamente 90 Mb a aproximadamente 100 Mb.

40 En otros casos, el inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser de al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 120 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos al menos 350 kb, al menos 400 kb, al menos 450 kb, al menos 500 kb, al menos 550 kb, al menos 600 kb, al menos 650 kb, al menos 700 kb o más.

45 El inserto de ácido nucleico puede comprender ADN genómico o cualquier otro tipo de ADN. Por ejemplo, el inserto de ácido nucleico puede ser de un procarionta, un eucariota, una levadura, un ave (p. ej., pollo), un mamífero no humano, un roedor, un ser humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un gato, un perro, un hurón, un primate (por ejemplo, tití, mono rhesus), un mamífero domesticado, un mamífero agrícola o cualquier otro organismo de interés.

50 El inserto de ácido nucleico y/o el ácido nucleico en el locus diana pueden comprender una secuencia codificante o una secuencia no codificante, tal como un elemento regulador (p. ej., un promotor, un potenciador o un elemento de unión a un represor transcripcional). Por ejemplo, el inserto de ácido nucleico puede comprender un alelo de inserción de al menos un exón de un gen endógeno, o un alelo de inserción del gen endógeno completo (es decir, "inserción de intercambio de genes"). Por ejemplo, el inserto de ácido nucleico puede ser homólogo u ortólogo a una secuencia a la que se dirige para su eliminación en el locus genómico diana. El inserto de ácido nucleico homólogo u ortólogo puede reemplazar la secuencia a la que se dirige la eliminación en el locus genómico de interés. Esto puede resultar en la humanización de un locus si la inserción del inserto de ácido nucleico da como resultado el reemplazo de una secuencia de ácido nucleico no humano con una secuencia de ácido nucleico humano homóloga u ortóloga (es decir, el inserto de ácido nucleico se inserta en lugar de la correspondiente secuencia de ADN humana en su locus genómico endógeno).

60 El inserto de ácido nucleico también puede comprender un alelo condicional. El alelo condicional puede ser un alelo multifuncional, como se describe en el documento US 2011/0104799. Por ejemplo, el alelo condicional puede comprender:



(a) una secuencia de actuación en orientación sentido con respecto a la transcripción de un gen diana;

5 (b) un casete de selección de fármacos (DSC) en orientación sentido o antisentido; (c) una secuencia de nucleótidos de interés (NSI) en orientación antisentido; y (d) un módulo condicional por inversión (COIN, que utiliza un intrón de división de exón y un módulo similar a una trampa de genes invertible) en orientación inversa. Véase, por ejemplo, el documento US 2011/0104799. El alelo condicional puede comprender además unidades recombinantes que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de activación y la DSC; y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido.  
10 Véase el documento US 2011/0104799.

Algunos insertos de ácido nucleico comprenden un polinucleótido que codifica un marcador de selección. El marcador de selección puede estar contenido en un casete de selección. Dichos marcadores de selección incluyen, pero no se limitan a, neomicina fosfotransferasa (*neo<sup>r</sup>*), higromicina B fosfotransferasa (*hyg<sup>r</sup>*), puromicina-N-acetiltransferasa (puro<sup>r</sup>), blasticidina S desaminasa (*bsr<sup>r</sup>*), xantina/guanina fosforribosil transferasa (*gpt*), o timidina quinasa del virus del herpes simple (HSV-k), o una combinación de las mismas. El polinucleótido que codifica el marcador de selección se puede unir operativamente a un promotor activo en una célula a la que se dirige. Los ejemplos de promotores se describen en otra parte del presente documento.

20 En algunos vectores de direccionamiento, el inserto de ácido nucleico comprende un gen informador. Ejemplos de genes informadores son genes que codifican luciferasa,  $\beta$ -galactosidasa, proteína fluorescente verde (GFP), proteína fluorescente verde mejorada (eGFP), proteína fluorescente cian (CFP), proteína fluorescente amarilla (YFP), proteína fluorescente amarilla mejorada (eYFP), proteína fluorescente azul (BFP), proteína fluorescente azul mejorada (eBFP), DsRed, ZsGreen, Mm GFP, mPlum, mCherry, tdTomato, m Strawberry, J-Red, mOrange, mKO, mCitrine, Venus, YPet, Emerald, CyPet, Cerulean, T-Sapphire, fosfatasa alcalina y una combinación de los mismos. Dichos genes informadores se pueden vincular operativamente a un promotor activo en una célula a la que se dirigen. Los ejemplos de promotores se describen en otra parte del presente documento.

30 En algunos vectores de direccionamiento, el inserto de ácido nucleico comprende uno o más casetes de expresión o casetes de eliminación. Un casete dado puede comprender una secuencia de nucleótidos de interés, un ácido nucleico que codifica un marcador de selección y/o un gen informador, junto con varios componentes reguladores que influyen en la expresión. Los ejemplos de marcadores seleccionables y genes informadores que pueden incluirse se discuten en detalle en otra parte del presente documento.

35 En algunos vectores de direccionamiento, el inserto de ácido nucleico comprende un ácido nucleico flanqueado por secuencias diana de recombinación específicas del sitio. Aunque el inserto de ácido nucleico completo puede estar flanqueado por tales secuencias diana de recombinación específicas de sitio, cualquier región o polinucleótido individual de interés dentro del inserto de ácido nucleico también puede estar flanqueado por dichos sitios. Las secuencias diana de recombinación específicas del sitio, que pueden flanquear el inserto de ácido nucleico o cualquier polinucleótido de interés en el inserto de ácido nucleico pueden incluir, por ejemplo, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox y una combinación de los mismos. En un ejemplo, los sitios de recombinación específicos del sitio flanquean un polinucleótido que codifica un marcador de selección y/o un gen informador contenido dentro del inserto de ácido nucleico. Después de la integración del inserto de ácido nucleico en un locus diana, se pueden eliminar las secuencias entre los sitios de recombinación específicos del sitio.

#### 45 B. Polinucleótidos de interés

Cualquier polinucleótido de interés puede estar contenido en los diversos insertos de ácido nucleico y, por lo tanto, integrado en el locus genómico diana. Los métodos divulgados en el presente documento prevén que al menos 1, 2, 50 3, 4, 5, 6 o más polinucleótidos de interés se integren en el locus genómico diana.

El polinucleótido de interés dentro del inserto de ácido nucleico cuando se integra en el locus genómico diana puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una eliminación de una secuencia de ácido nucleico endógeno y/o la adición de un polinucleótido exógeno, heterólogo u ortólogo en el locus genómico diana. En una realización, la modificación genética comprende la sustitución de una secuencia de ácido nucleico endógeno por un polinucleótido exógeno de interés en el locus genómico diana. Por lo tanto, los métodos proporcionados en este documento permiten la generación de una modificación genética que comprende una inactivación, una eliminación, una inserción, un reemplazo ("introducción"), una mutación puntual, un cambio de dominio, un cambio de exón, un cambio de intrón, un cambio de secuencias reguladoras, un cambio de genes o una combinación de los mismos en un locus genómico diana. Dichas modificaciones pueden ocurrir tras la integración de la primera, segunda, tercera, cuarta, quinta, sexta, séptima o cualquier inserto de ácido nucleico subsiguiente en el locus genómico diana.

65 El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana puede comprender una secuencia que es nativa u homóloga a la célula en la que se introduce; el polinucleótido de interés puede ser heterólogo a la célula en la que se introduce; el polinucleótido de interés puede ser exógeno a la célula en

la que se introduce; el polinucleótido de interés puede ser ortólogo a la célula en la que se introduce; o el polinucleótido de interés puede ser de una especie diferente a la célula en la que se introduce. "Homólogo" en referencia a una secuencia incluye una secuencia que es nativa de la célula. "Heterólogo" en referencia a una secuencia incluye una secuencia que se origina a partir de una especie foránea o, si es de la misma especie, se modifica sustancialmente de su forma nativa en composición y/o locus genómico por intervención humana deliberada. "Exógeno" en referencia a una secuencia incluye una secuencia que se origina a partir de una especie foránea. "Ortólogo" incluye un polinucleótido de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie (es decir, una variante de especie). El polinucleótido de interés puede ser de cualquier organismo de interés, incluidos, pero sin limitarse a, un no humano, un roedor, un hámster, un ratón, una rata, un humano, un mono, un ave, un mamífero agrícola o un mamífero no agrícola. El polinucleótido de interés puede comprender además una región codificante, una región no codificante, una región reguladora o un ADN genómico. Por lo tanto, los insertos de ácido nucleico 1°, 2°, 3°, 4°, 5°, 6°, 7° y/o cualquiera de los subsiguientes insertos de ácido nucleico pueden comprender dichas secuencias.

En una realización, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana es homólogo a un ácido nucleico humano. En aún otras realizaciones, el polinucleótido de interés integrado en el locus diana es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mamífero agrícola o un ácido nucleico de mamífero no agrícola o una combinación de los mismos.

En una realización, el polinucleótido de interés puede oscilar de aproximadamente 500 nucleótidos a aproximadamente 200 kb como se describió anteriormente. El polinucleótido de interés puede ser de aproximadamente 500 nucleótidos a aproximadamente 5 kb, de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 700 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 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 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, de aproximadamente 190 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb o de aproximadamente 600 kb a aproximadamente 700 kb.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico diana puede codificar un polipéptido, puede codificar un miARN, puede codificar un ARN largo no codificante o puede comprender cualquier región reguladora o región no codificante de interés que incluyen, por ejemplo, una secuencia reguladora, una secuencia promotora, una secuencia potenciadora, una secuencia de unión a un represor transcripcional o una eliminación de una secuencia que no codifica una proteína, pero no comprende una eliminación de una secuencia que codifica proteína. Además, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico diana puede codificar una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo, o una combinación de los mismos.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana puede comprender una modificación genética en una secuencia codificante. Dichas modificaciones genéticas incluyen, pero no se limitan a, una mutación por eliminación de una secuencia codificante o la fusión de dos secuencias codificantes.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana puede comprender un polinucleótido que codifica una proteína mutante. En una realización, la proteína mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado. En una realización, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico comprende al menos un alelo de enfermedad. En tales casos, el alelo de la enfermedad puede ser un alelo dominante o el alelo de la enfermedad es un alelo recesivo. Además, el alelo de la enfermedad puede comprender un alelo de polimorfismo de un solo nucleótido (SNP). El polinucleótido de interés que codifica la proteína mutante puede ser de cualquier organismo, incluidos, pero sin limitarse a, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un ser humano, un mono, un mamífero agrícola o un polinucleótido de mamífero doméstico que codifica una proteína mutante.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana también puede comprender una secuencia reguladora, que incluye, por ejemplo, una secuencia promotora, una secuencia potenciadora, una secuencia de unión a un represor transcripcional o una secuencia terminadora transcripcional. En realizaciones específicas, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus

genómico diana comprende un polinucleótido que tiene una eliminación de una secuencia que no codifica una proteína, pero no comprende una eliminación de una secuencia que codifica una proteína. En una realización, la eliminación de la secuencia que no codifica la proteína comprende la eliminación de una secuencia reguladora. En otra realización, la eliminación del elemento regulador comprende la eliminación de una secuencia promotora. En una realización, la eliminación del elemento regulador comprende la eliminación de una secuencia potenciadora. Dicho polinucleótido de interés puede ser de cualquier organismo, incluidos, pero sin limitarse a, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un ser humano, un mono, un mamífero agrícola o un polinucleótido de mamífero doméstico que codifica una proteína mutante.

Una modificación genética dirigida puede comprender una alteración dirigida a un polinucleótido de interés. Dichas modificaciones dirigidas incluyen, pero sin limitarse a, adiciones de uno o más nucleótidos, eliminaciones de uno o más nucleótidos, sustituciones de uno o más nucleótidos, una desactivación del polinucleótido de interés o una porción del mismo, una inserción del polinucleótido de interés o una porción del mismo, un reemplazo de una secuencia de ácido nucleico endógena con una secuencia de ácido nucleico heteróloga, o una combinación de los mismos. En realizaciones específicas, al menos 1, 2, 3, 4, 5, 7, 8, 9, 10, 100, 500 o más nucleótidos o al menos 10 kb a 500 kb o más se cambian para formar la modificación genómica dirigida.

### C. Vectores de direccionamiento

Los vectores de direccionamiento pueden emplearse para introducir el inserto de ácido nucleico en un locus diana genómico y comprenden el inserto de ácido nucleico y los brazos de homología que flanquean el inserto de ácido nucleico. Los vectores de direccionamiento pueden estar en forma lineal o en forma circular, y pueden ser monocatenarios o bicatenarios. Los vectores de direccionamiento pueden ser ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). Para facilitar la referencia, los brazos de homología se denominan en este documento brazos de homología 5' y 3' (es decir, secuencia arriba y secuencia abajo). Esta terminología se relaciona con la posición relativa de los brazos de homología con respecto al inserto de ácido nucleico dentro del vector de direccionamiento. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus diana o a una región dentro de otro vector de direccionamiento, que se denominan en el presente documento "secuencia diana 5'" y "secuencia diana 3'", respectivamente. En algunos casos, los brazos de homología también pueden funcionar como una secuencia diana 5' o 3'.

Los presentes métodos emplean dos, tres o más vectores de direccionamiento que son capaces de recombinarse entre sí. En varias realizaciones, los vectores de direccionamiento son vectores de direccionamiento grandes (LTVEC) como se describe en otra parte del presente documento. En tales métodos, el primer, el segundo y el tercer vector de direccionamiento comprenden cada uno un brazo de homología 5' y 3'. El brazo de homología 3' del primer vector de direccionamiento comprende una secuencia que se superpone con el brazo de homología 5' del segundo vector de direccionamiento (es decir, secuencias superpuestas), lo que permite la recombinación homóloga entre el primero y el segundo LTVEC.

En el caso de métodos de doble direccionamiento, el brazo de homología 5' del primer vector de direccionamiento y el brazo de homología 3' del segundo vector de direccionamiento son homólogos a segmentos correspondientes dentro del locus genómico diana (es decir, la secuencia diana) que promueve la recombinación homóloga del primer y el segundo vector de direccionamiento con los segmentos genómicos correspondientes y modifica el locus genómico diana.

En el caso de métodos de direccionamiento triple, el brazo de homología 3' del segundo vector de direccionamiento comprende una secuencia que se superpone con el brazo de homología 5' del tercer vector de direccionamiento (es decir, secuencias superpuestas), lo que permite la recombinación homóloga entre la segunda y la tercera LTVEC. El brazo de homología 5' del primer vector de direccionamiento y el brazo de homología 3' del tercer vector de direccionamiento son homólogos a los segmentos correspondientes dentro del locus genómico diana (es decir, la secuencia diana) que promueve la recombinación homóloga del primer y tercer vector de direccionamiento con los segmentos genómicos correspondientes y modifica el locus genómico diana.

Un brazo de homología y una secuencia diana o dos brazos de homología "se corresponden" o "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. 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 el vector de direccionamiento (es decir, la secuencia superpuesta) o entre dos brazos de homología 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 del vector de direccionamiento (o un fragmento del mismo) y la secuencia diana de otro vector de direccionamiento o una secuencia diana del locus genómico diana (o un fragmento del mismo) puede tener 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 modo 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 en el locus genómico diana. Por ejemplo, un brazo de homología dado y/o la secuencia diana correspondiente pueden comprender regiones correspondientes de homología que tienen al menos aproximadamente 5-10 kb, 5-15 kb, 5-20 kb, 5-25 kb, 5-30 kb, 5-35 kb, 5-40 kb, 5-45 kb, 5-50 kb, 5-55 kb, 5-60 kb, 5-65 kb, 5-70 kb, 5-75 kb, 5-80 kb, 5-85 kb, 5-90 kb, 5-95 kb, 5-100 kb, 100-200 kb o 200-300 kb de longitud o más (tal como se describe en los vectores LTVEC descritos en otra parte del presente documento) de manera que el brazo de homología tiene homología suficiente para sufrir una recombinación homóloga con las secuencias diana correspondientes dentro del locus genómico diana de la célula o dentro de otro vector de direccionamiento.

Las secuencias superpuestas del brazo de homología 3' del primer vector de direccionamiento y el brazo de homología 5' del segundo vector de direccionamiento o del brazo de homología 3' del segundo vector de direccionamiento y el brazo de homología 5' del tercer vector de direccionamiento pueden tener cualquier longitud que sea suficiente para promover la recombinación homóloga entre los vectores de direccionamiento. Por ejemplo, una secuencia superpuesta dada de un brazo de homología puede comprender regiones superpuestas correspondientes que tienen al menos aproximadamente 1-5 kb, 5-10 kb, 5-15 kb, 5-20 kb, 5-25 kb, 5-30 kb, 5-35 kb, 5-40 kb, 5-45 kb, 5-50 kb, 5-55 kb, 5-60 kb, 5-65 kb, 5-70 kb, 5-75 kb, 5-80 kb, de 5 a 85 kb, de 5 a 90 kb, de 5 a 95 kb, de 5 a 100 kb, de 100 a 200 kb o de 200 a 300 kb de longitud o más, de modo que la secuencia superpuesta del brazo de homología tenga homología suficiente para experimentar recombinación homóloga con la correspondiente secuencia superpuesta dentro de otro vector de direccionamiento. En una realización, la secuencia superpuesta es de 1 a 5 kb. En una realización, la secuencia superpuesta es de aproximadamente 1 kb a aproximadamente 70 kb. En una realización, la secuencia superpuesta es de aproximadamente 10 kb a aproximadamente 70 kb. En otra realización, la secuencia superpuesta es de aproximadamente 10 kb a aproximadamente 50 kb. En una realización, la secuencia superpuesta es de al menos 10 kb. En otra realización, la secuencia superpuesta es de al menos 20 kb. Por ejemplo, la secuencia superpuesta puede ser de aproximadamente 1 kb a aproximadamente 5 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 15 kb, de aproximadamente 15 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 25 kb, de aproximadamente 25 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 35 kb, de aproximadamente 35 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 45 kb, de aproximadamente 45 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 220 kb, de aproximadamente 220 kb a aproximadamente 240 kb, de aproximadamente 240 kb a aproximadamente 260 kb, de aproximadamente 260 kb a aproximadamente 280 kb, o de aproximadamente 280 kb a aproximadamente 300 kb. Como ejemplo, la secuencia superpuesta puede ser de aproximadamente 20 kb a aproximadamente 60 kb. Alternativamente, la secuencia superpuesta puede ser de al menos 1 kb, al menos 5 kb, al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 25 kb, al menos 30 kb, al menos 35 kb, al menos 40 kb, al menos 45 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 120 kb, al menos 140 kb, al menos 160 kb, al menos 180 kb, al menos 200 kb, al menos 220 kb, al menos 240 kb, al menos 260 kb, al menos 280 kb, o al menos 300 kb.

Los brazos de homología pueden corresponder a un locus nativo de una célula (p. ej., el locus diana) o, alternativamente, pueden corresponder a una región de un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula incluidos, por ejemplo, transgenes, casetes de expresión o regiones heterólogas o exógenas de ADN. Alternativamente, los brazos de homología pueden corresponder a una región en un vector de direccionamiento en una célula. Los brazos de homología del vector de direccionamiento pueden corresponder a una región de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región diseñada contenida en una célula huésped apropiada. Aún más, los brazos de homología del vector de direccionamiento pueden corresponder o derivarse de una región de una biblioteca BAC, una biblioteca de cósmidos o una biblioteca de fagos PI. En ciertos casos, los brazos de homología del vector de direccionamiento corresponden a un locus que es nativo, heterólogo o exógeno a un procarionte, una levadura, un ave (p. ej., pollo), un mamífero no humano, un roedor, un ser humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un gato, un perro, un hurón, un primate (p. ej., tití, mono rhesus), un mamífero domesticado, un mamífero agrícola, o cualquier otro organismo de interés. En algunos casos, los brazos de homología corresponden a un locus de la célula que no se puede seleccionar con un método convencional o que se puede seleccionar de manera incorrecta o con una eficiencia significativamente baja en ausencia de un corte o rotura de doble cadena inducida por un agente de nucleasa (por ejemplo, una proteína Cas). En algunos casos, los brazos de homología se derivan de ADN sintético.

En algunos vectores de direccionamiento, uno de los brazos de homología 5' o 3' corresponde a un locus genómico diana mientras que el otro de los brazos de homología 5' o 3' corresponde a una región en otro vector de direccionamiento.

En algunos vectores de direccionamiento, los brazos de homología 5' y 3' corresponden a un genoma dirigido. Alternativamente, los brazos de homología pueden ser de un genoma relacionado. Por ejemplo, el genoma diana es un genoma de ratón de una primera cepa, y los brazos dirigidos son de un genoma de ratón de una segunda cepa, en la que la primera cepa y la segunda cepa son diferentes. En ciertos casos, los brazos de homología son del genoma del mismo animal o son del genoma de la misma cepa, por ejemplo, el genoma diana es un genoma de ratón de una primera cepa, y los brazos de direccionamiento son de un genoma de ratón del mismo ratón o de la misma cepa.

Un brazo de homología de un vector de direccionamiento puede tener cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con una secuencia diana correspondiente, que incluye, por ejemplo, al menos 1-5 kb, 5-10 kb, 5-15 kb, 5-20 kb, 5-25 kb, 5-30 kb, 5-35 kb, 5-40 kb, 5-45 kb, 5-50 kb, 5-55 kb, 5-60 kb, 5-65 kb, 5-70 kb, 5-75 kb, 5-80 kb, 5-85 kb, 5-90 kb, 5-95 kb, 5-100 kb, 100-200 kb o 200-300 kb de longitud o mayor. Como se describe con mayor detalle a continuación, los vectores de direccionamiento grandes pueden emplear brazos de direccionamiento de mayor longitud.

Los agentes de nucleasa (p. ej., sistemas CRISPR/Cas) se pueden emplear en combinación con vectores de direccionamiento para ayudar en la modificación de un locus diana. Dichos agentes de nucleasa pueden promover la recombinación homóloga entre el vector de direccionamiento y el locus diana. Cuando los agentes de nucleasa se emplean en combinación con un vector de direccionamiento, el vector de direccionamiento puede comprender brazos de homología 5' y 3' correspondientes a secuencias diana 5' y 3' ubicadas lo suficientemente cerca de un sitio de escisión de nucleasa para promover la aparición de un evento de recombinación homóloga entre las secuencias diana y los brazos de homología tras un corte o rotura de doble cadena en el sitio de escisión de la nucleasa. El término "sitio de escisión de nucleasa" incluye una secuencia de ADN en la que un agente de nucleasa (por ejemplo, un sitio de escisión Cas9) crea un corte o rotura de cadena doble. Las secuencias diana dentro del locus diana que corresponden a los brazos de homología 5' y 3' del vector de direccionamiento están "ubicadas lo suficientemente cerca" de un sitio de escisión de nucleasa si la distancia es tal que promueve la aparición de un evento de recombinación homóloga entre las secuencias diana 5' y 3' y los brazos de homología tras un corte o rotura de doble cadena en el sitio de reconocimiento. Por lo tanto, en casos específicos, las secuencias diana correspondientes a los brazos de homología 5' y/o 3' del vector de direccionamiento están dentro de al menos 1 nucleótido de un sitio de reconocimiento dado o están dentro de al menos 10 nucleótidos a aproximadamente 14 kb de un sitio de reconocimiento dado. En algunos casos, el sitio de escisión de la nucleasa está inmediatamente adyacente a al menos una o ambas secuencias diana.

La relación espacial de las secuencias diana que corresponden a los brazos de homología del vector de direccionamiento y el sitio de escisión de la nucleasa puede variar. Por ejemplo, las secuencias diana se pueden ubicar 5' con respecto al sitio de escisión de la nucleasa, las secuencias diana se pueden ubicar 3' con respecto al sitio de reconocimiento, o las secuencias diana pueden flanquear el sitio de escisión de la nucleasa.

El uso combinado del vector de direccionamiento (incluyendo, por ejemplo, un vector de direccionamiento grande) con un agente de nucleasa puede dar como resultado una eficiencia de direccionamiento aumentada en comparación con el uso del vector de direccionamiento solo. Por ejemplo, cuando un vector de direccionamiento se usa junto con un agente de nucleasa, la eficiencia de direccionamiento del vector de direccionamiento puede incrementarse al menos dos veces, al menos tres veces, al menos 4 veces, al menos 10 veces o dentro de un intervalo formado por estos números enteros, tal como de 2 a 10 veces en comparación con el uso del vector de direccionamiento solo.

#### D. Grandes vectores de direccionamiento

Los vectores de direccionamiento son "vectores de direccionamiento grandes" o "LTVEC", que incluyen vectores de direccionamiento que comprenden brazos de homología que corresponden y se derivan de secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar recombinación homóloga en células. De acuerdo con los métodos de la invención, el LTVEC tiene al menos 10 kb de longitud, o la suma total del brazo de homología 5' y el brazo de homología 3' puede ser, por ejemplo, al menos 10 kb. Los LTVEC también incluyen vectores de direccionamiento que comprenden insertos de ácido nucleico que tienen secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en las células. Por ejemplo, los LTVEC hacen posible la modificación de grandes loci que no pueden adaptarse a los vectores tradicionales basados en plásmidos debido a sus limitaciones de tamaño. Por ejemplo, el locus diana puede ser (es decir, los brazos de homología 5' y 3' pueden corresponder a) un locus de la célula que no se puede direccionar usando un método convencional o que se puede direccionar solamente incorrectamente o solo con una eficiencia significativamente baja en la ausencia de un corte o rotura de doble cadena inducida por un agente de nucleasa (p. ej., una proteína Cas).

Los métodos proporcionados en el presente documento emplean dos o tres LTVEC que son capaces de recombinarse entre sí y con el locus genómico diana en un evento de recombinación de tres o cuatro vías como se describe en otra parte del presente documento. Estos métodos hacen posible la modificación de grandes loci que no se pueden lograr utilizando un solo LTVEC.

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 de LTVEC y métodos para fabricarlos, por ejemplo, en la patente de los Estados Unidos N° 6,586,251; Patente de Estados Unidos N° 6,596,541; Patente de Estados Unidos N° 7,105,348; y el documento WO 2002/036789 (PCT/USOI/45375). Los LTVEC pueden ser de forma lineal o de forma circular.

Los LTVEC pueden tener cualquier longitud, incluidos, por ejemplo, de aproximadamente 20 kb a aproximadamente 300 kb, de aproximadamente 20 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. Alternativamente, un LTVEC 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 mayor. El tamaño de un LTVEC puede ser demasiado grande para permitir la detección de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de largo alcance (por ejemplo, de 1 kb a 5 kb).

En algunos casos, un LTVEC comprende un inserto de ácido nucleico que varía de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 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 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb. En otros casos, el inserto de ácido nucleico puede variar 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. En algunos casos, un LTVEC comprende un inserto de ácido nucleico que varía de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 650 kb, de aproximadamente 650 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 750 kb, o de aproximadamente 750 kb a aproximadamente 800 kb.

En algunos LTVECS, la suma total del brazo de homología 5' y el brazo de homología 3' es de al menos 10 kb. En otros LTVEC, el brazo de homología 5' varía de aproximadamente 1 kb a aproximadamente 100 kb y/o el brazo de homología 3' varía de aproximadamente 1 kb a aproximadamente 100 kb. La suma total de los brazos de homología 5' y 3' puede ser, por ejemplo, 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 de 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 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb. Alternativamente, cada brazo de homología puede tener al menos 5 kb, 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 110 kb, al menos 120 kb, al menos 130 kb, al menos 140 kb, al menos 150 kb, al menos 160 kb, al menos 170 kb, al menos 180 kb, al menos 190 kb o al menos 200 kb. Asimismo, la suma total de los brazos de homología 5' y 3' puede ser de al menos 5 kb, 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 110 kb, al menos 120 kb, al menos 130 kb, al menos 140 kb, al menos 150 kb, al menos 160 kb, al menos 170 kb, al menos 180 kb, al menos 190 kb o al menos 200 kb.

En algunos casos, el LTVEC y el inserto de ácido nucleico están diseñados para permitir una eliminación de la secuencia endógena en el locus diana 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, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 800 kb, de aproximadamente 800 kb a aproximadamente 1 Mb, de aproximadamente 1 Mb a aproximadamente 1.5 Mb, de aproximadamente 1.5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2.5 Mb, o de aproximadamente 2.5 Mb a aproximadamente 3 Mb . Alternativamente, la eliminación puede ser de aproximadamente 3 Mb a aproximadamente 4 Mb, de aproximadamente 4 Mb a aproximadamente 5 Mb, de aproximadamente 5 Mb a aproximadamente 10 Mb, de aproximadamente 10 Mb a aproximadamente 20 Mb, de aproximadamente 20 Mb a aproximadamente 30 Mb, de aproximadamente 30 Mb a aproximadamente 40 Mb, de aproximadamente 40 Mb a aproximadamente 50 Mb, de aproximadamente 50 Mb a aproximadamente 60 Mb, de aproximadamente 60 Mb a aproximadamente 70 Mb, de aproximadamente 70 Mb a aproximadamente 80 Mb, de aproximadamente 80 Mb a aproximadamente 90 Mb, o de aproximadamente 90 Mb a aproximadamente 100 Mb. Alternativamente, la eliminación puede ser de al menos 10 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 otros casos, el LTVEC y el inserto de ácido nucleico están diseñados para permitir una inserción en el locus diana de una secuencia de ácido nucleico exógeno que varía 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, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. Alternativamente, la inserción puede ser de aproximadamente 400 kb a aproximadamente 450 kb, de aproximadamente 450 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 550 kb, de aproximadamente 550 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 650 kb, de aproximadamente 650 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 750 kb, o de aproximadamente 750 kb a aproximadamente 800 kb. Alternativamente, la inserción puede ser de al menos 10 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 otros casos más, el inserto de ácido nucleico y/o la región del locus endógeno que se está eliminando tiene al menos 100, 200, 300, 400, 500, 600, 700, 800 o 900 nucleótidos o al menos 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb o más.

#### E. Genomas y loci genómicos diana

Un genoma o locus diana genómico modificado por los métodos divulgados en el presente documento puede incluir cualquier segmento o región de ADN dentro de una célula. El genoma o locus diana genómico puede ser nativo de la célula, puede ser un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula, o puede ser una combinación de los mismos. Dichos segmentos heterólogos o exógenos de ADN pueden incluir transgenes, casetes de expresión, marcadores de selección que codifican polinucleótidos o regiones heterólogas o exógenas de ADN genómico.

El genoma o locus diana genómico también puede incluir ADN extracromosómico dentro de la célula, tal como un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región genómica modificada contenida en una célula huésped apropiada.

#### III. Agentes de nucleasa

Los métodos y composiciones para modificar un locus genómico diana proporcionados en el presente documento emplean un agente de nucleasa que induce un corte o rotura de doble cadena en un sitio de reconocimiento deseado.

El término "sitio de reconocimiento para un agente de nucleasa" incluye una secuencia de ADN en la que se induce un corte o rotura de doble cadena mediante un agente de nucleasa. El sitio de reconocimiento para un agente de nucleasa puede ser endógeno (o nativo) a la célula o el sitio de reconocimiento puede ser exógeno a la célula. En realizaciones específicas, el sitio de reconocimiento es exógeno a la célula y, por lo tanto, no se encuentra de forma natural en el genoma de la célula. En aún otras realizaciones, el sitio de reconocimiento es exógeno a la célula y a los polinucleótidos de interés que se desea que se posicionen en el locus diana. En realizaciones adicionales, el sitio de reconocimiento exógeno o endógeno está presente solo una vez en el genoma de la célula huésped. En realizaciones específicas, se identifica un sitio endógeno o nativo que ocurre solo una vez dentro del genoma. Dicho sitio se puede

usar para diseñar agentes de nucleasa que producirán un corte o una rotura de doble cadena en el sitio de reconocimiento endógeno.

5 La longitud del sitio de reconocimiento puede variar e incluye, por ejemplo, sitios de reconocimiento 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 similar a un activador de transcripción (TALEN), o aproximadamente 20 pb para un ARN guía de CRISPR/Cas9.

10 Cualquier agente de nucleasa que induce un corte o rotura de doble cadena en un sitio de reconocimiento deseado puede usarse en los métodos y composiciones divulgados en el presente documento. Puede emplearse un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca un corte o rotura de doble cadena en un sitio de reconocimiento deseado. Alternativamente, se puede emplear un agente de nucleasa modificado o manipulado. Un "agente de nucleasa modificado" incluye una nucleasa que está modificada (modificada o derivada) a partir de su forma nativa para reconocer e inducir específicamente un corte o rotura de doble cadena en el sitio de reconocimiento deseado. Por lo tanto, un agente de nucleasa modificado puede derivarse de un agente de nucleasa nativo que se produce de forma natural o puede crearse o sintetizarse artificialmente. La modificación del agente de 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. En algunas realizaciones, la nucleasa modificada induce un corte o rotura de doble cadena en un sitio de reconocimiento, en el que el sitio de reconocimiento no era una secuencia que habría sido reconocida por un agente de nucleasa nativo (no modificado o no manipulado). La producción de un corte o rotura de doble cadena en un sitio de reconocimiento u otro ADN puede denominarse en el presente documento "cortar" o "escindir" el sitio de reconocimiento u otro ADN.

25 También se proporcionan variantes activas y fragmentos de los sitios de reconocimiento ejemplificados. 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 el sitio de reconocimiento dado, en el que las variantes activas conservan la actividad biológica y, por lo tanto, pueden ser reconocidas y escindidas por un agente de nucleasa de una manera específica de secuencia. Se conocen en la técnica ensayos para medir la rotura de doble cadena de un sitio de reconocimiento por un agente de nucleasa (p. ej., ensayo qPCR TAQMAN®, Frenthewey D. et al., *Methods in Enzymology*, 2010, 476:295-307).

35 El sitio de reconocimiento del agente de nucleasa se puede colocar en cualquier lugar dentro o cerca del locus diana. El sitio de reconocimiento puede estar ubicado dentro de una región codificante de un gen o dentro de regiones reguladoras que influyen en la expresión del gen. Un sitio de reconocimiento del agente de nucleasa se puede ubicar en un intrón, un exón, un promotor, un potenciador, una región reguladora o cualquier región codificante que no sea de proteína. En realizaciones específicas, el sitio de reconocimiento se coloca dentro del polinucleótido que codifica el marcador de selección. Dicha posición puede ubicarse dentro de la región codificante del marcador de selección o dentro de las regiones reguladoras, que influyen en la expresión del marcador de selección. Por lo tanto, un sitio de reconocimiento del agente de nucleasa se puede ubicar en un intrón del marcador de selección, un promotor, un potenciador, una región reguladora o cualquier región que no codifica proteínas del polinucleótido que codifica el marcador de selección. En realizaciones específicas, un corte o rotura de doble cadena en el sitio de reconocimiento interrumpe la actividad del marcador de selección. Se conocen métodos para ensayar la presencia o ausencia de un marcador de selección funcional.

45 En una realización, el agente de nucleasa es una nucleasa efectora similar a un activador de transcripción (TALEN). Las nucleasas efectoras TAL son una clase de nucleasas específicas de secuencia que se pueden usar para hacer roturas de doble cadena en secuencias diana específicas en el genoma de un organismo procarionta o eucariota. Las nucleasas efectoras TAL se crean fusionando un efector similar a un activador de la transcripción (TAL) nativo o modificado, o una parte funcional del mismo, con el dominio catalítico de una endonucleasa, como, por ejemplo, FokI. El dominio de unión al ADN efector TAL modular único permite el diseño de proteínas con potencialmente cualquier especificidad de reconocimiento de ADN dada. Por lo tanto, los dominios de unión al ADN de las nucleasas efectoras TAL pueden modificarse para reconocer sitios diana de ADN específicos y, por lo tanto, usarse para hacer roturas de doble cadena en las secuencias diana deseadas. Véase el documento WO 2010/079430; Morbitzer et al. (2010) PNAS 10.1073/pnas.1013133107; Scholze & Boch (2010) *Virulence* 1 :428-432; Christian et al. *Genetics* (2010) 186:757-761; Li et al. (2010) *Nuc. Acids Res.* (2010) doi: 10.1093/nar/gkq704; y Miller et al., (2011) *Nature Biotechnology* 29: 143-148.

60 Se describen ejemplos de nucleasas TAL adecuadas y métodos para preparar nucleasas TAL adecuadas, por ejemplo, en los documentos US 2011/0239315 AI, US 2011/0269234 AI, US 2011/0145940 AI, US 2003/0232410 AI, US 2005/0208489 AI, US 2005/0026157 AI, US 2005/0064474 AI, US 2006/0188987 AI y US 2006/0063231 AI. En diversas realizaciones, las nucleasas efectoras TAL están modificadas para cortar en o cerca de una secuencia de ácido nucleico diana en, por ejemplo, un locus de interés o un locus de interés genómico, en el que la secuencia de ácido nucleico diana está en o cerca de una secuencia que va a ser modificada por un vector de direccionamiento. Las nucleasas TAL adecuadas para su uso con los diversos métodos y composiciones proporcionados en el presente documento incluyen aquellas que están específicamente diseñadas para unirse a secuencias de ácido nucleico diana o cerca de ellas para ser modificadas por vectores de direccionamiento como se describe en el presente documento.

65



En una realización, cada monómero de TALEN comprende 33-35 repeticiones TAL que reconocen un solo par de bases a través de dos residuos hipervariables. En una realización, el agente de nucleasa es una proteína quimérica que comprende un dominio de unión a ADN basado en repeticiones TAL unido operativamente a una nucleasa independiente. En una realización, la nucleasa independiente es una endonucleasa FokI. En una realización, el agente de nucleasa comprende un primer dominio de unión a ADN basado en repeticiones TAL y un segundo dominio de unión a ADN basado en repeticiones TAL, en el que cada uno del primer y segundo dominios de unión a ADN basados en repeticiones TAL está unido operativamente a una nucleasa FokI, en la que el primer y el segundo dominio de unión a ADN basado en la repetición 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 la que las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que hace que una cadena doble se rompa en una secuencia diana.

El agente de nucleasa empleado en los diversos métodos y composiciones divulgados en el presente documento puede comprender además una nucleasa con dedos de zinc (ZFN). En una realización, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de zinc, en el que cada dominio de unión a ADN basado en dedos de zinc se une a un subsitio de 3 pb. En otras realizaciones, 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. En una realización, la endonucleasa independiente es una endonucleasa FokI. En una realización, el agente de nucleasa comprende un primer ZFN y un segundo ZFN, en el que cada uno del primer ZFN y el segundo ZFN está unido operativamente a una subunidad de nucleasa FokI, en el que el primer y el segundo 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 el que las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que hace que se rompa una doble cadena. Véase, por ejemplo, los documentos US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293 A2; y Gaj et al. (2013) Trends in Biotechnology, 31(7):397- 405.

En aún otra realización, el agente de nucleasa es una meganucleasa. Las meganucleasas se han clasificado en cuatro familias basadas en motivos de secuencias conservadas, las familias son las familias de caja LAGLIDADG, GIY-YIG, H-N-H y His-Cys. Estos motivos participan en la coordinación de iones metálicos y la hidrólisis de enlaces fosfodiéster. Las meganucleasas se destacan por sus largos sitios de reconocimiento y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Se conocen los dominios, la estructura y la función de la meganucleasa, véase, por ejemplo, Guhan y Muniyappa (2003) Crit Rev Biochem Mol Biol 38: 199-248; Lucas et al., (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 et al., (2002) Nat Struct Biol 9:764. En algunos ejemplos, se usa una variante de origen natural y/o un derivado modificado de meganucleasa. Se conocen métodos para modificar la cinética, las interacciones de cofactores, la expresión, las condiciones óptimas y/o la especificidad del sitio de reconocimiento y la detección de actividad, véase, por ejemplo, Epinat et al., (2003) Nucleic Acids Res 31 :2952-62; Chevalier et al., (2002) Mol Cell 10:895-905; Gimble et al., (2003) Mol Biol 334:993-1008; Seligman et al., (2002) Nucleic Acids Res 30:3870-9; Sussman et al., (2004) J Mol Biol 342:31-41; Rosen et al., (2006) Nucleic Acids Res 34:4791-800; Chames et al., (2005) Nucleic Acids Res 33 :e178; Smith et al., (2006) Nucleic Acids Res 34:e149; Gruen et al., (2002) Nucleic Acids Res 30:e29; Chen y Zhao, (2005) Nucleic Acids Res 33 :e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.

Cualquier meganucleasa se puede utilizar en este documento, incluyendo, pero sin limitarse a, I-Scel, I-Scell, I-ScellII, I-ScelIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I -CeuAIIP, I-Crel, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PspI, F-Scel, F-Scell, F-SuvI, F-TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I-CvuAIP, I-Ddil, I-Ddill, I -DirI, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-Llal, I-Msol, I-Naal, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-Njal, I-Nsp236IP, I-PakI, I-PboIP, I-PculP, I-PcuAl, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I -Scal, I-SexIP, I-SnelP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbilP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-Pful, PI- Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-Scel, PI-Tful, PI-Tfull, PI-Thyl, PI-TIil, PI-THII o cualquiera de sus variantes activas o fragmentos.

En una realización, la meganucleasa reconoce secuencias de ADN de doble cadena de 12 a 40 pares de bases. En una realización, la meganucleasa reconoce una secuencia diana perfectamente coincidente en el genoma. En una realización, la meganucleasa es una nucleasa dirigida. En una realización, la nucleasa dirigida es una familia LAGLIDADG de nucleasas dirigidas. En una realización, la familia LAGLIDADG de nucleasas dirigidas se selecciona de I-Scel, I-Crel e I-Dmol.

Los agentes de nucleasa pueden comprender además 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 normalmente 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 normalmente 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 Ila 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 IIs reconocen un sitio de reconocimiento asimétrico y escinden por un lado y a 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 et al., (2003) *Nucleic Acids Res* 31 :418-20), Roberts et al., (2003) *Nucleic Acids Res* 31: 1805-12, y Belfort et al., (2002) en *Mobile DNA II*, págs. 761-783, Eds. Craigie et al., (ASM Press, Washington, DC).

El agente de nucleasa empleado en los diversos métodos y composiciones también puede comprender un sistema de repeticiones palindrómicas cortas agrupadas regularmente intercaladas (CRISPR)/asociado a CRISPR (Cas) o componentes de tal sistema. Los sistemas CRISPR/Cas incluyen transcripciones y otros elementos involucrados en la expresión o direccionamiento de la actividad de los genes Cas. Un sistema CRISPR/Cas puede ser un sistema de tipo I, tipo II o tipo III. Los métodos y composiciones divulgados en el presente documento emplean sistemas CRISPR/Cas mediante la utilización de complejos CRISPR (que comprenden un ARN guía (ARNg) complejado con una proteína Cas) para la escisión de ácidos nucleicos dirigida al sitio.

Algunos sistemas CRISPR/Cas utilizados en los métodos divulgados en el presente documento no se producen de forma natural. Un sistema de "ocurrencia 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 natural, estando al menos sustancialmente libre de al menos otro componente con que están naturalmente asociados en la naturaleza, o que están asociados con al menos otro componente con el que no están naturalmente asociados. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR no naturales que comprenden un ARNg y una proteína Cas que no se presentan juntos de forma natural.

Las proteínas Cas generalmente comprenden al menos un dominio de unión o reconocimiento de ARN. Dichos dominios pueden interactuar con los ARN guía (ARNg, que se describen con más detalle a continuación). Las proteínas Cas también pueden comprender dominios de nucleasa (por ejemplo, dominios de DNasa o RNasa), dominios de unión a ADN, dominios de helicasa, dominios de interacción proteína-proteína, dominios de dimerización y otros dominios. Un dominio de nucleasa posee actividad catalítica para la escisión de ácidos nucleicos. La escisión 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 monocatenario o bicatenario. Una proteína Cas puede tener una actividad de escisión completa y crear una rotura de doble cadena en un locus genómico diana (p. ej., una rotura de doble cadena con extremos romos), o puede ser una nickasa que crea una rotura de cadena sencilla en un locus genómico diana.

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 versiones homólogas o modificadas de las mismas.

En algunos casos, una proteína Cas es de un sistema CRISPR/Cas de tipo II. Por ejemplo, la proteína Cas puede ser una proteína Cas9 o derivarse de una proteína Cas9. Las proteínas Cas9 suelen compartir 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. La proteína Cas9 puede ser de, por ejemplo, *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Nocardiaopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycooides*, *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 beccii*, *Candidatus Desulfuridis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochrochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia sp. umigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lynngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus* o *Acaryochloris marina*. Las proteínas Cas9 también pueden ser de *Staphylococcus aureus*. Los ejemplos adicionales de los miembros de la familia Cas9 incluyen los descritos en el documento WO 2014/131833. En un ejemplo específico, la proteína Cas9 es una proteína Cas9 de *S. pyogenes* o se deriva de ella. La secuencia de aminoácidos de una proteína Cas9 de *S. pyogenes* se puede encontrar, por ejemplo, en la base de datos SwissProt con el número de acceso Q99ZW2.

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 activas o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las variantes o fragmentos activos pueden comprender al menos 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 la que las variantes activas conservan la capacidad de cortar en un sitio de escisión deseado y, por lo tanto, conservan la actividad inductora de cortes o inductora de rotura de doble cadena. Se conocen ensayos para la actividad inductora de cortes o inductora de rotura de doble cadena y, en general, miden la actividad global y la especificidad de la proteína Cas en sustratos de ADN que contienen el sitio de escisión.

Las proteínas Cas se pueden modificar para aumentar o disminuir la afinidad de unión de ácidos nucleicos, la especificidad de unión de ácidos nucleicos y/o la actividad enzimática. Las proteínas Cas también se pueden modificar 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 se pueden modificar, eliminar o inactivar, o una proteína Cas se puede trincar para eliminar dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad de la proteína Cas.

Algunas proteínas Cas comprenden al menos dos dominios de nucleasa, tales como los dominios de DNasa. Por ejemplo, una proteína Cas9 puede comprender 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. Véase, por ejemplo, Jinek et al. (2012) *Science* 337:816-821.

Uno o ambos dominios de nucleasa se pueden eliminar o mutar para que ya no sean funcionales o tengan una actividad de nucleasa reducida. Si uno de los dominios de la nucleasa se elimina o muta, la proteína Cas resultante (p. ej., Cas9) puede denominarse nickasa y puede generar una ruptura de una sola cadena en una secuencia diana dentro de un ADN de doble cadena, pero no una ruptura de doble cadena (es decir, puede escindir la cadena complementaria o la cadena no complementaria, pero no ambas). Si ambos dominios de nucleasa se eliminan o mutan, la proteína Cas resultante (p. ej., Cas9) tendrá una capacidad reducida para escindir ambas cadenas de un ADN de doble cadena (p. ej., una proteína Cas sin nucleasa). Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato por alanina en la posición 10 de Cas9) en el dominio RuvC de Cas9 de *S. pyogenes*. Asimismo, H939A (histidina por alanina en la posición del aminoácido 839) o H840A (histidina por alanina en la posición del aminoácido 840) en el dominio HNH de Cas9 de *S. pyogenes* puede convertir Cas9 en una nickasa. Otros ejemplos de mutaciones que convierten a Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Véase, por ejemplo, Sapranaukas et al. (2011) *Nucleic Acids Research* 39:9275-9282 y el documento WO 2013/141680. Dichas mutaciones se pueden generar utilizando métodos bien conocidos, tales como la mutagénesis dirigida al sitio, la mutagénesis mediada por PCR o la síntesis total de genes. Se pueden encontrar ejemplos de otras mutaciones que crean nickasas, por ejemplo, en los documentos WO/2013/176772A1 y WO/2013/142578A1.

Las proteínas Cas también pueden ser proteínas de fusión. Por ejemplo, una proteína Cas se puede fusionar con un dominio de escisión, un dominio de modificación epigenética, un dominio de activación transcripcional o un dominio represor transcripcional. Véase el documento WO 2014/089290. Las proteínas Cas también se pueden fusionar con un polipéptido heterólogo proporcionando una mayor o menor estabilidad. El dominio fusionado o el polipéptido heterólogo puede ubicarse en el extremo terminal N, el extremo terminal C o internamente dentro de la proteína Cas.

Un ejemplo de una proteína de fusión Cas es una proteína Cas fusionada con un polipéptido heterólogo que proporciona la localización subcelular. Dichas secuencias pueden incluir, por ejemplo, una señal de localización nuclear (NLS) tal como NLS de SV40 para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a las mitocondrias, una señal de retención ER y similares. Véase, por ejemplo, Lange et al. (2007) *J. Biol. Chem.* 282:5101-5105. Una proteína Cas puede comprender, por ejemplo, una o más señales de localización nuclear (por ejemplo, dos señales de localización nuclear). Dichas señales de localización subcelular pueden ubicarse en el extremo terminal N, el extremo terminal C o en cualquier lugar dentro de la proteína Cas. Un NLS puede comprender un tramo de aminoácidos básicos y puede ser una secuencia monopartita o una secuencia bipartita.

Las proteínas Cas también pueden comprender un dominio de penetración celular. Por ejemplo, el dominio de penetración celular puede derivarse 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 del Herpes simple o una secuencia peptídica de poliarginina. Véase, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular se puede ubicar en el extremo terminal N, el extremo terminal C o en cualquier lugar dentro de la proteína Cas.

Las proteínas Cas también pueden comprender un polipéptido heterólogo para facilitar el seguimiento o la purificación, tal como una proteína fluorescente, una etiqueta de purificación o una etiqueta de epítipo. 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, ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), proteínas fluorescentes azules (p. ej., eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), proteínas fluorescentes cian (p. ej., eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas fluorescentes rojas (mKate, mKate2, mPlum, DsRed-Monómero, mCherry, mRFPI, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tándem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), proteínas fluorescentes naranjas (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) y cualquier otra proteína fluorescente adecuada. Los ejemplos de etiquetas incluyen glutatión-S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tioredoxina (TRX), poli(NANP), etiqueta de

purificación por afinidad en tándem (TAP), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemaglutinina (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidina (His), proteína portadora de biotina carboxilo (BCCP) y calmodulina.

5 Las proteínas Cas se pueden proporcionar en cualquier forma. Por ejemplo, una proteína Cas se puede proporcionar en forma de una proteína, tal como una proteína Cas complejada con un ARNg. Alternativamente, una proteína Cas se puede proporcionar en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (p. ej., ARN mensajero (ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas puede tener codones optimizados para una traducción eficiente en proteína en una célula u organismo particular. Cuando se introduce en la célula un ácido nucleico que codifica la proteína Cas, la proteína Cas puede expresarse de manera transitoria, condicional o constitutiva en la célula.

10 Los ácidos nucleicos que codifican proteínas Cas pueden integrarse de forma estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican las proteínas Cas pueden unirse operativamente a un promotor en una construcción de expresión.

15 Las construcciones de expresión incluyen cualquier construcción de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (p. ej., un gen Cas) y que puede transferir tal secuencia de ácido nucleico de interés a una célula diana. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede estar en un vector que comprende un ADN que codifica un ARNg. Alternativamente, puede estar en un vector o plásmido que esté separado del vector que comprende el ADN que codifica el ARNg. Los promotores que se pueden usar en una construcción de expresión incluyen, por ejemplo, promotores activos en una célula pluripotente de rata, eucariota, de mamífero, de mamífero no humano, de humano, roedor, ratón o hámster. Los ejemplos de otros promotores se describen en otra parte del presente documento.

20 Un "ARN guía" o "ARNg" incluye una molécula de ARN que se une a una proteína Cas y dirige la proteína Cas a una ubicación 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 un segmento, sección o región de una molécula, tal como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg comprenden dos moléculas de ARN separadas: un "ARN activador" y un "ARN direccionador". Otros ARNg son una sola molécula de ARN (polinucleótido de ARN único), que también se puede denominar "ARNg de molécula única", un "ARN de guía única" o un "ARNgu". Véase, por ejemplo, los documentos WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622 A2, WO/2014/099750A2, WO/2013142578A1 y WO 2014/131833A1. Los términos "ARN guía" y "ARNg" son inclusivos e incluyen tanto ARNg de molécula doble como ARNg de molécula sencilla.

25 Un ejemplo de ARNg de dos moléculas comprende una molécula similar a ARNcr ("ARN CRISPR" o "ARN direccionador" o "ARNcr" o "repetición de ARNcr") y una molécula similar a ARNtracr correspondiente ("ARN CRISPR que actúa en forma trans" o "ARN activador" o "ARNtracr" o "estructura"). Un ARNcr comprende tanto el segmento de direccionamiento al ADN (monocatenario) del ARNg y un tramo de nucleótidos que forma la mitad del dúplex de ARNbc del segmento de unión a proteínas del ARNg.

30 Un ARNtracr correspondiente (ARN activador) comprende un tramo de nucleótidos que forma la otra mitad del dúplex de ARNbc del segmento de unión a proteínas del ARNg. Un tramo de nucleótidos de un ARNcr es complementario e hibrida con un tramo de nucleótidos de un ARNtracr para formar el dúplex de ARNbc del dominio de unión a proteínas del ARNg. Como tal, se puede decir que cada ARNcr tiene un ARNtracr correspondiente.

35 El ARNcr y el ARNtracr correspondiente se hibridan para formar un ARNg. El ARNcr proporciona además el segmento de direccionamiento al ADN monocatenario que se hibrida con una secuencia diana. Si se utiliza 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 utilizarán las moléculas de ARN. Véase, por ejemplo, Malí et al. (2013) Science 339:823-826; Jinek et al. (2012) Science 337:816-821; Hwang et al. (2013) Nat. Biotechnol. 31 :227-229; Jiang et al. (2013) Nat. Biotechnol. 31 :233-239; y Cong et al. (2013) Science 339:819-823.

40 El segmento de direccionamiento al ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN diana. El segmento de direccionamiento al ADN de un ARNg interactúa con un ADN diana de una manera específica de secuencia mediante hibridación (es decir, emparejamiento de bases). Como tal, la secuencia de nucleótidos del segmento de direccionamiento al ADN puede variar y determina la ubicación dentro del ADN diana con el que interactuarán el ARNg y el ADN diana. El segmento de direccionamiento al ADN de un ARNg dado se puede modificar para hibridar con cualquier secuencia deseada dentro de un ADN diana. Los ARNcr de origen natural difieren según el sistema Cas9 y el organismo, pero a menudo 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 (véase, por ejemplo, el documento WO2014/131833). En el caso de *S. pyogenes*, los DR tienen una longitud de 36 nucleótidos y el segmento de direccionamiento tiene una longitud de 30 nucleótidos. El DR localizado en 3' es complementario e hibrida con el ARNtracr correspondiente, que a su vez se une a la proteína Cas9.

El segmento de direccionamiento al ADN puede tener una longitud de aproximadamente 12 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, el segmento de direccionamiento al ADN puede tener una longitud de aproximadamente 12 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 12 nt a aproximadamente 50 nt, de aproximadamente 12 nt a aproximadamente 40 nt, de aproximadamente 12 nt a aproximadamente 30 nt, de aproximadamente 12 nt a aproximadamente 25 nt, de aproximadamente 12 nt a aproximadamente 20 nt, o de aproximadamente 12 nt a aproximadamente 19 nt. Alternativamente, el segmento de direccionamiento al ADN puede tener una longitud de aproximadamente 19 nt a aproximadamente 20 nt, de aproximadamente 19 nt a aproximadamente 25 nt, de aproximadamente 19 nt a aproximadamente 30 nt, de aproximadamente 19 nt a aproximadamente 35 nt, de aproximadamente 19 nt a aproximadamente 40 nt, de aproximadamente 19 nt a aproximadamente 45 nt, de aproximadamente 19 nt a aproximadamente 50 nt, de aproximadamente 19 nt a aproximadamente 60 nt, de aproximadamente 19 nt a aproximadamente 70 nt, de aproximadamente 19 nt a aproximadamente 80 nt, de aproximadamente 19 nt a aproximadamente 90 nt, de aproximadamente 19 nt a aproximadamente 100 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 20 nt a aproximadamente 35 nt, de aproximadamente 20 nt a aproximadamente 40 nt, de aproximadamente 20 nt a aproximadamente 45 nt, de aproximadamente 20 nt a aproximadamente 50 nt, de aproximadamente 20 nt a aproximadamente 60 nt, de aproximadamente 20 nt a aproximadamente 70 nt, de aproximadamente 20 nt a aproximadamente 80 nt, de aproximadamente 20 nt a aproximadamente 90 nt, o de aproximadamente 20 nt a aproximadamente 100 nt.

La secuencia de nucleótidos del segmento de direccionamiento al ADN que es complementaria a una secuencia de nucleótidos (secuencia diana) del ADN diana puede tener una longitud de al menos aproximadamente 12 nt. Por ejemplo, la secuencia de direccionamiento al ADN (es decir, la secuencia dentro del segmento de direccionamiento al ADN que es complementaria a una secuencia diana dentro del ADN diana) puede tener una longitud de al menos aproximadamente 12 nt, al menos aproximadamente 15 nt, al menos aproximadamente 18 nt, al menos aproximadamente 19 nt, al menos aproximadamente 20 nt, al menos aproximadamente 25 nt, al menos aproximadamente 30 nt, al menos aproximadamente 35 nt, o al menos aproximadamente 40 nt. Alternativamente, la secuencia de direccionamiento al ADN puede tener una longitud de aproximadamente 12 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 12 nt a aproximadamente 50 nt, de aproximadamente 12 nt a aproximadamente 45 nt, de aproximadamente 12 nt a aproximadamente 40 nt, de aproximadamente 12 nt a aproximadamente 35 nt, de aproximadamente 12 nt a aproximadamente 30 nt, de aproximadamente 12 nt a aproximadamente 25 nt, de aproximadamente 12 nt a aproximadamente 20 nt, de aproximadamente 12 nt a aproximadamente 19 nt, de aproximadamente 19 nt a aproximadamente 20 nt, de aproximadamente 19 nt a aproximadamente 25 nt, de aproximadamente 19 nt a aproximadamente 30 nt, de aproximadamente 19 nt a aproximadamente 35 nt, de aproximadamente 19 nt a aproximadamente 40 nt, de aproximadamente 19 nt a aproximadamente 45 nt, de aproximadamente 19 nt a aproximadamente 50 nt, de aproximadamente 19 nt a aproximadamente 60 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 20 nt a aproximadamente 35 nt, de aproximadamente 20 nt a aproximadamente 40 nt, de aproximadamente 20 nt a aproximadamente 45 nt, de aproximadamente 20 nt a aproximadamente 50 nt, o de aproximadamente 20 nt a aproximadamente 60 nt. En algunos casos, la secuencia de direccionamiento al ADN puede tener una longitud de aproximadamente 20 nt.

Los ARNtracr pueden estar en cualquier forma (p. ej., ARNtracr de longitud completa o ARNtracr parciales activos) y de diferentes longitudes. Pueden incluir transcripciones primarias o formas procesadas. 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 o consistir en la totalidad o una parte de una secuencia de ARNtracr de tipo silvestre (p. ej., aproximadamente o más de aproximadamente 20, 26, 32, 45, 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. Véase, por ejemplo, Deltcheva et al. (2011) Nature 471: 602-607; documento WO 2014/093661. Los ejemplos de ARNtracr dentro de los ARN de guía única (ARNgu) incluyen los segmentos de ARNtracr que se encuentran dentro de las versiones +48, +54, +67 y +85 de los ARNgu, donde "+n" indica que hasta el nucleótido +n de ARNtracr de tipo silvestre está incluido en el ARNgu. Véase el documento US 8,697,359.

El porcentaje de complementariedad entre la secuencia de direccionamiento al ADN y la secuencia diana dentro del ADN diana puede ser de al menos el 60 % (por ejemplo, al menos el 65 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 %). En algunos casos, el porcentaje de complementariedad entre la secuencia de direccionamiento al ADN y la secuencia diana dentro del ADN diana es de al menos el 60 % sobre aproximadamente 20 nucleótidos contiguos. En un ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento al ADN y la secuencia diana dentro del ADN diana es del 100 % sobre los 14 nucleótidos contiguos en el extremo 5' de la secuencia diana dentro de la cadena complementaria del ADN diana y tan bajo como 0 % sobre el resto. En tal caso, se puede considerar que la secuencia de direccionamiento al ADN tiene una longitud de 14 nucleótidos. En otro ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento al ADN y la secuencia diana dentro del ADN diana es del 100 % sobre los siete nucleótidos contiguos en el extremo 5' de la secuencia diana dentro de la cadena complementaria del ADN diana y tan bajo como 0 % sobre el resto. En tal caso, se puede considerar que la secuencia de direccionamiento al ADN tiene una longitud de 7 nucleótidos.

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 (ARNbc). El segmento de unión a proteínas de un ARNg dado 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.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (p. ej., estabilidad modificada o regulada; direccionamiento subcelular; seguimiento 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 casquete 5' (por ejemplo, un casquete de 7-metilguanilato (m7G)); una cola poliadenilada 3' (es decir, una cola poli(A) 3'); una secuencia de riboconmutador (p. ej., 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 ARNbc (es decir, una horquilla); una modificación o secuencia que dirige el ARN a una ubicación subcelular (p. ej., núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona seguimiento (p. ej., conjugación directa con una molécula fluorescente, conjugación con una fracción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.); una modificación o secuencia que proporciona un sitio de unión para proteínas (p. ej., proteínas que actúan sobre el ADN, incluidos activadores transcripcionales, represores transcripcionales, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares); y combinaciones de los mismos.

Los ARN guía se pueden proporcionar en cualquier forma. Por ejemplo, el ARNg se puede proporcionar en forma de ARN, ya sea como dos moléculas (ARNcr y ARNtracr separados) o como una molécula (ARNgu), y opcionalmente en forma de un complejo con una proteína Cas. El ARNg también se puede proporcionar en forma de ADN que codifica el ARNg. El ADN que codifica el ARNg puede codificar una sola molécula de ARN (ARNgu) o moléculas de ARN separadas (p. ej., ARNcr y ARNtracr separados). En el último caso, el ADN que codifica el ARNg se puede proporcionar como moléculas de ADN separadas que codifican el ARNcr y el ARNtracr, respectivamente. Alternativamente, el ADN que codifica el ARNg se puede proporcionar como una molécula de ADN.

Cuando un ADN que codifica un ARNg se introduce en una célula, el ARNg puede expresarse de forma transitoria, condicional o constitutiva en la célula. Los ADN que codifican los ARNg pueden integrarse de forma estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg se pueden unir operativamente a un promotor en una construcción de expresión. Por ejemplo, el ADN que codifica el ARNg puede estar en un vector que comprende un ácido nucleico que codifica una proteína Cas. Alternativamente, puede estar en un vector o un plásmido que esté separado del vector que comprende el ácido nucleico que codifica la proteína Cas. Los promotores que se pueden usar en dichas construcciones de expresión incluyen promotores activos, por ejemplo, en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. En algunos casos, el promotor es un promotor de ARN polimerasa III, tal como un promotor U6 humano.

Alternativamente, los ARNg se pueden preparar mediante varios otros métodos. Por ejemplo, los ARNg se pueden preparar mediante transcripción *in vitro* usando, por ejemplo, ARN polimerasa T7 (véase, por ejemplo, los documentos WO 2014/089290 y WO 2014/065596). Los ARN guía también pueden ser una molécula producida sintéticamente preparada por síntesis química.

Una secuencia diana para un sistema CRISPR/Cas incluye secuencias de ácido nucleico presentes en un ADN diana al que se unirá un segmento de direccionamiento al ADN de un ARNg, siempre que existan condiciones suficientes para la unión. Por ejemplo, las secuencias diana incluyen secuencias para las que se diseña un ARN guía para que tenga complementariedad, donde la hibridación entre una secuencia diana y una secuencia de direccionamiento al ADN 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.

Las secuencias diana también incluyen sitios de escisión para proteínas Cas, que se describen con más detalle a continuación. Una secuencia diana puede comprender cualquier polinucleótido, que puede ubicarse, por ejemplo, en el núcleo o el citoplasma de una célula o dentro de un orgánulo de una célula, tal como una mitocondria o un cloroplasto.

La secuencia diana dentro de un ADN diana puede estar dirigida por (es decir, unirse, hibridarse o ser complementaria a) una proteína Cas o un ARNg. 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 (p. ej., condiciones en un sistema libre de células) (véase, p. ej., *Molecular Cloning: A Laboratory Manual*, 3ª edición (Sambrook et al., Harbor Laboratory Press 2001)). La cadena del ADN diana que es complementaria y se hibrida a la proteína Cas o ARNg puede denominarse la "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) se puede llamar "cadena no complementaria" o "cadena plantilla".

La proteína Cas puede escindir el ácido nucleico en un sitio dentro o fuera de una secuencia de ácido nucleico presente en un ADN diana al que se unirá un segmento de direccionamiento al ADN de un ARNg. El "sitio de escisión" incluye la posición de un ácido nucleico en el que una proteína Cas produce una rotura de una sola cadena o una rotura de doble cadena. Por ejemplo, la formación de un complejo CRISPR (que comprende un ARNg hibridado con una secuencia diana y complejo con una proteína Cas) puede resultar en la escisión de una o ambas cadenas en o cerca (p. ej., dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases de) la secuencia de ácido nucleico presente en un ADN diana al que se unirá un segmento de direccionamiento al ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico presente en un ADN diana al que se unirá un segmento de direccionamiento al ADN de un ARNg, se sigue considerando que el sitio de escisión está dentro de la "secuencia diana". El sitio de escisión puede estar en una sola cadena o en ambas cadenas de un ácido nucleico. Los sitios de escisión pueden estar en la misma posición en ambas cadenas del ácido nucleico (produciendo extremos romos) o pueden estar en diferentes sitios en cada cadena (produciendo extremos escalonados (es decir, salientes)). Los extremos escalonados se pueden producir, por ejemplo, utilizando dos proteínas Cas que producen una rotura de una sola cadena en diferentes sitios de escisión en diferentes cadenas, produciendo así una rotura de doble cadena. Por ejemplo, una primera nickasa puede crear una rotura de una sola cadena en la primera cadena de ADN bicatenario (ADNbc), mientras que una segunda nickasa puede crear una rotura de una sola cadena en la segunda cadena de ADNbc de modo que se creen secuencias salientes. En algunos casos, la secuencia diana de la nickasa en la primera cadena está separada de la secuencia diana 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 1,000 pares de bases.

La escisión específica del sitio del ADN diana por parte de Cas9 puede ocurrir en ubicaciones determinadas por (i) la complementariedad de emparejamiento de bases entre el ARNg y el ADN diana y (ii) un motivo corto, denominado motivo adyacente al protoespaciador (PAM), en el ADN diana. El PAM puede flanquear la secuencia diana. Opcionalmente, la secuencia diana puede estar flanqueada en el extremo 3' por el PAM. Por ejemplo, el sitio de escisión de Cas9 puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) secuencia arriba o secuencia abajo de la secuencia PAM. En algunos casos (p. ej., cuando se usa Cas9 de *S. pyogenes* o un Cas9 estrechamente relacionado), la secuencia PAM de la cadena no complementaria puede ser 5'-N<sub>1</sub>GG-3', donde N<sub>1</sub> es cualquier nucleótido de ADN y es inmediatamente 3' de la secuencia diana de la cadena no complementaria del ADN diana. Como tal, la secuencia PAM de la cadena complementaria sería 5'-CCN<sub>2</sub>-3', donde N<sub>2</sub> es cualquier nucleótido de ADN y está inmediatamente 5' de la secuencia diana de 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 (p. ej., 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).

Los ejemplos de secuencias diana incluyen una secuencia de ADN complementaria al segmento de direccionamiento al ADN de un ARNg, o tal secuencia de ADN además de una secuencia PAM. Un ejemplo de una secuencia diana comprende la secuencia de nucleótidos de GNNNNNNNNNNNNNNNNNNNNNGG (GN<sub>1-20</sub>GG; SEQ ID NO: 1). La guanina en el extremo 5' puede facilitar la transcripción por la ARN polimerasa en las células. Otros ejemplos de secuencias diana pueden incluir dos nucleótidos de guanina en el extremo 5' para facilitar la transcripción eficiente por polimerasa T7 *in vitro*. Véase, por ejemplo, el documento WO 2014/065596. Otras secuencias diana pueden tener entre 4 y 22 nucleótidos de longitud de la SEQ ID NO: 1, incluidos 5' G y 3' GG. Otras secuencias diana más pueden tener entre 14 y 20 nucleótidos de longitud de la SEQ ID NO: 1.

La secuencia diana puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia diana puede ser una secuencia que codifica un producto génico (p. ej., una proteína) o una secuencia no codificante (p. ej., una secuencia reguladora o ADN basura) o puede incluir ambas.

También se proporcionan variantes y fragmentos activos de agentes de nucleasa (es decir, un agente de nucleasa modificado). 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 el agente de nucleasa nativo, en el que las variantes activas retienen la capacidad de cortar en un sitio de reconocimiento deseado y, por lo tanto, retienen la actividad inductora de corte o rotura de doble cadena. Por ejemplo, cualquiera de los agentes de nucleasa descritos en el presente documento puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir un corte o rotura de doble cadena en un sitio de reconocimiento que no fue reconocido por el agente de nucleasa nativo. Por lo tanto, en algunas realizaciones, la nucleasa modificada tiene una especificidad para inducir un corte o rotura de doble cadena en un sitio de reconocimiento que es diferente del sitio de reconocimiento del agente de nucleasa nativo correspondiente. Se conocen ensayos para actividad de corte o inductora de rotura de doble cadena y generalmente miden la actividad global y la especificidad de la endonucleasa en sustratos de ADN que contienen el sitio de reconocimiento.

El agente de nucleasa se puede introducir en la célula pluripotente por cualquier medio conocido en la técnica. El polipéptido que codifica el agente de nucleasa se puede introducir directamente en la célula. Alternativamente, se puede introducir en la célula un polinucleótido que codifica el agente de nucleasa. Cuando se introduce en la célula

un polinucleótido que codifica el agente de nucleasa, el agente de nucleasa puede expresarse de manera transitoria, condicional o constitutiva dentro de la célula. Por lo tanto, el polinucleótido que codifica el agente de nucleasa puede estar contenido en un casete de expresión y estar unido operativamente a un promotor condicional, un promotor inducible, un promotor constitutivo o un promotor específico de tejido. Alternativamente, el agente de nucleasa se introduce en la célula como un ARNm que codifica un agente de nucleasa.

En realizaciones específicas, el polinucleótido que codifica el agente de nucleasa se integra de forma estable en el genoma de la célula y se une operativamente a un promotor activo en la célula. En otras realizaciones, el polinucleótido que codifica el agente de nucleasa está en el mismo vector de direccionamiento que comprende el inserto de ácido nucleico, mientras que en otros casos el polinucleótido que codifica el agente de nucleasa está en un vector o plásmido que está separado del vector de direccionamiento que comprende el inserto de ácido nucleico.

Cuando el agente de nucleasa se proporciona a la célula a través de la introducción de un polinucleótido que codifica el agente de nucleasa, dicho polinucleótido que codifica un agente de nucleasa se puede modificar para sustituir los codones que tienen una frecuencia de uso más alta en la célula de interés, en comparación con la secuencia de polinucleótidos de origen natural que codifica el agente de nucleasa. Por ejemplo, el polinucleótido que codifica el agente de nucleasa se puede modificar para sustituir los codones que tienen una mayor frecuencia de uso en una célula procarionota o eucariota, incluida 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 polinucleótidos de origen natural.

Los diversos métodos expuestos anteriormente pueden repetirse secuencialmente para permitir la integración dirigida de cualquier número de insertos de ácido nucleico en un locus genómico diana dado en un cromosoma. Por lo tanto, los diversos métodos prevén la inserción de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más insertos de ácido nucleico en el locus genómico diana en un cromosoma. En realizaciones particulares, dichos métodos de mosaico secuencial permiten la reconstrucción de grandes regiones genómicas de una célula animal o de una célula de mamífero (es decir, un humano, un no humano, un roedor, un ratón, un mono, una rata, un hámster, un mamífero domesticado o un animal agrícola) en un locus genómico diana en un cromosoma. En tales casos, la transferencia y reconstrucción de regiones genómicas que incluyen tanto regiones codificantes como no codificantes permiten preservar la complejidad de una región determinada conservando, al menos en parte, las regiones codificantes, las regiones no codificantes y las variaciones del número de copias encontradas dentro de la región genómica nativa. Así, los diversos métodos proporcionan, por ejemplo, métodos para generar regiones genómicas "heterólogas" o "exógenas" dentro de una célula.

#### IV. Marcadores de selección

Los diversos métodos y composiciones proporcionados en este documento pueden emplear los agentes de nucleasa y sus sitios de reconocimiento correspondientes en combinación con marcadores de selección. Como se analiza en el presente documento, la posición del sitio de reconocimiento en el polinucleótido que codifica el marcador de selección permite un método eficaz mediante el cual identificar los eventos de integración en el locus diana. Además, en el presente documento se proporcionan varios métodos en los que se emplean marcadores de selección alternos que tienen el sitio de reconocimiento de nucleasa para mejorar la eficiencia y la eficacia a través de las cuales se integran múltiples polinucleótidos de interés dentro de un locus diana dado.

Se pueden usar varios marcadores de selección en los métodos y composiciones divulgados en el presente documento. Dichos marcadores de selección pueden, por ejemplo, impartir resistencia a un antibiótico tal como G418, higromicina, blastidina, neomicina o puromicina. Dichos marcadores de selección incluyen neomicina fosfotransferasa (*neo<sup>r</sup>*), higromicina B fosfotransferasa (*hyg<sup>r</sup>*), puromicina-N-acetiltransferasa (*puro<sup>r</sup>*) y blastidina S desaminasa (*bsr<sup>r</sup>*). En aún otras realizaciones, el marcador de selección está unido operativamente a un promotor inducible y la expresión del marcador de selección es tóxica para la célula. Los ejemplos no limitantes de tales marcadores de selección incluyen xantina/guanina fosforribosil transferasa (*gpt*), hipoxantina-guanina fosforribosiltransferasa (*HGPRT*) o timidina quinasa del virus del herpes simple (*HSV-TK*).

En una realización, el sitio de reconocimiento de nucleasa se coloca dentro de un gen que codifica un marcador de selección. En una realización específica, el sitio de reconocimiento de la nucleasa está posicionado dentro de un gen de higromicina.

El polinucleótido que codifica los marcadores de selección está operativamente unido a un promotor activo en la célula. Dichos casetes de expresión y sus diversos componentes reguladores se discuten con mayor detalle en otra parte del presente documento.

#### V. Promotores

Varias secuencias de ácido nucleico descritas en el presente documento pueden unirse operativamente a los promotores. Dichos promotores pueden ser activos, por ejemplo, en una célula pluripotente, eucariota, de mamífero, de mamífero no humano, humana, de roedor, de ratón o de hámster. Un promotor puede ser, por ejemplo, un promotor



constitutivamente activo, un promotor condicional, un promotor inducible, un promotor temporalmente restringido (p. ej., un promotor regulado por el desarrollo) o un promotor espacialmente restringido (p. ej., un promotor específico de células o específico de tejidos). Se pueden encontrar ejemplos de promotores, por ejemplo, en el documento WO 2013/176772.

5 Los ejemplos de promotores inducibles incluyen, por ejemplo, promotores químicamente regulados y promotores físicamente regulados. Los promotores regulados químicamente incluyen, por ejemplo, promotores regulados por alcohol (p. ej., un promotor del gen de alcohol deshidrogenasa (alcA)), promotores regulados por tetraciclina (p. ej., un promotor sensible a tetraciclina, una secuencia operadora de tetraciclina (tetO), un promotor tet-On o un promotor tet-Off), promotores regulados por esteroides (p. ej., un receptor de glucocorticoides de rata, un promotor de un receptor de estrógeno o un promotor de un receptor de ecdisona) o promotores regulados por metales (p. ej., un promotor de metaloproteína). Los promotores regulados físicamente incluyen, por ejemplo, promotores regulados por temperatura (por ejemplo, un promotor de choque térmico) y promotores regulados por luz (por ejemplo, un promotor inducible por luz o un promotor reprimible por luz).

15 Los promotores específicos de tejido pueden ser, por ejemplo, promotores específicos de neuronas, promotores específicos de glía, promotores específicos de células musculares, promotores específicos de células cardíacas, promotores específicos de células renales, promotores específicos de células óseas, promotores específicos de células endoteliales o promotores específicos de células inmunitarias (por ejemplo, un promotor de células B o un promotor de células T).

Los promotores regulados por el desarrollo incluyen, por ejemplo, promotores activos solo durante una etapa embrionaria de desarrollo, o solo en una célula adulta.

25 También se puede seleccionar un promotor basándose en el tipo de célula. Por ejemplo, varios promotores conocidos encuentran uso en una célula eucariota, una célula de mamífero, una célula no humana, una célula de mamífero no humano, una célula pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida en el desarrollo, una célula iPS humana, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster, un fibroblasto o una célula CHO.

## VI. Casetes de expresión

35 En este documento se proporcionan polinucleótidos o moléculas de ácido nucleico que comprenden los diversos componentes del sistema de direccionamiento proporcionado en este documento (es decir, agentes de nucleasa, sitios de reconocimiento, insertos de ácido nucleico, polinucleótidos de interés, vectores de direccionamiento (es decir, LTVEC), marcadores de selección y otros componentes).

40 Se proporcionan además polinucleótidos recombinantes que comprenden los diversos componentes del sistema de direccionamiento. Los términos "polinucleótido recombinante" y "construcción de ADN recombinante" se usan indistintamente en este documento. Una construcción recombinante comprende una combinación artificial o heteróloga de secuencias de ácidos nucleicos, por ejemplo, secuencias reguladoras y codificantes que no se encuentran juntas en la naturaleza. En otras realizaciones, una construcción recombinante puede comprender secuencias reguladoras y secuencias codificantes que se derivan de diferentes fuentes, o secuencias reguladoras y secuencias codificantes derivadas de la misma fuente, pero dispuestas de una manera diferente a la que se encuentra en la naturaleza. Tal construcción se puede usar sola o se puede usar junto con un vector. Si se usa un vector, entonces la elección del vector depende del método que se use para transformar las células huésped, como bien saben los expertos en la técnica. Por ejemplo, puede usarse un vector de plásmido. Los elementos genéticos necesarios para transformar, seleccionar y propagar con éxito las células huésped y que comprenden cualquiera de los fragmentos de ácido nucleico aislados se proporcionan en el presente documento. El cribado puede realizarse mediante análisis Southern de ADN, análisis Northern de expresión de ARNm, análisis de inmunotransferencia de expresión de proteínas o análisis fenotípico, entre otros.

55 En realizaciones específicas, uno o más de los componentes del sistema de direccionamiento descrito en el presente documento se pueden proporcionar en un casete de expresión para la expresión en una célula procariota, una célula eucariota, una bacteria, una célula de levadura, una célula de mamífero u otro organismo o tipo de célula de interés. El casete puede incluir secuencias reguladoras 5' y 3' unidas operativamente a un polinucleótido proporcionado en el presente documento. Cuando se usa para referirse a la unión de dos regiones codificantes de proteínas, ligadas operativamente significa que las regiones codificantes están en el mismo marco de lectura. En otro caso, una secuencia de ácido nucleico que codifica una proteína puede unirse operativamente a secuencias reguladoras (p. ej., secuencia promotora, potenciadora, silenciadora, etc.) para conservar la regulación transcripcional adecuada.

65 El casete puede contener adicionalmente al menos un polinucleótido adicional de interés para ser introducido conjuntamente en el organismo. Alternativamente, el polinucleótido de interés adicional se puede proporcionar en múltiples casetes de expresión. Dicho casete de expresión está provisto de una pluralidad de sitios de restricción y/o sitios de recombinación para la inserción de un polinucleótido recombinante que se encuentra bajo la regulación

transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes marcadores de selección.

5 El casete de expresión puede incluir en la dirección de transcripción 5'-3', una región de iniciación transcripcional y traduccional (es decir, un promotor), un polinucleótido recombinante proporcionado en el presente documento, y una  
 10 región de terminación transcripcional y traduccional (es decir, región de terminación) funcional en una célula de mamífero o una célula huésped de interés. Las regiones reguladoras (es decir, promotores, regiones reguladoras de la transcripción y regiones de terminación de la traducción) y/o un polinucleótido proporcionado en el presente  
 15 documento pueden ser nativos/análogos a la célula huésped o entre sí. Alternativamente, las regiones reguladoras y/o un polinucleótido proporcionado en el presente documento pueden ser heterólogos con respecto a la célula huésped o entre sí. Por ejemplo, un promotor unido operativamente a un polinucleótido heterólogo es de una especie diferente de la especie de la que se derivó el polinucleótido o, si es de la misma/análoga especie, uno o ambos están sustancialmente modificados a partir de su forma y/o locus originales, o el promotor no es el promotor nativo del polinucleótido unido operativamente. Alternativamente, las regiones reguladoras y/o un polinucleótido recombinante proporcionado en el presente documento pueden ser completamente sintéticos.

20 La región de terminación puede ser nativa con la región de inicio de la transcripción, puede ser nativa con el polinucleótido recombinante unido operativamente, puede ser nativa con la célula huésped o puede derivarse de otra fuente (es decir, extraña o heteróloga) al promotor, el polinucleótido recombinante, la célula huésped o cualquier combinación de los mismos.

25 Al preparar el casete de expresión, los diversos fragmentos de ADN pueden manipularse para proporcionar las secuencias de ADN en la orientación adecuada. Con este fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN o pueden estar implicadas otras manipulaciones para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción, o similares. Para este fin, pueden estar implicadas la mutagénesis *in vitro*, la reparación de cebadores, la restricción, la hibridación, las nuevas sustituciones, por ejemplo, transiciones y transversiones.

30 Se pueden usar varios promotores en los casetes de expresión proporcionados en el presente documento. Los promotores se pueden seleccionar en función del resultado deseado. Se reconoce que pueden mejorarse diferentes aplicaciones mediante el uso de diferentes promotores en los casetes de expresión para modular el tiempo, la ubicación y/o el nivel de expresión del polinucleótido de interés. Dichas construcciones de expresión también pueden contener, si se desea, una región reguladora del promotor (p. ej., una que confiere expresión inducible, constitutiva, regulada por el medio ambiente o el desarrollo, o específica/selectiva de células o tejidos), un sitio de inicio de la  
 35 transcripción, un sitio de unión al ribosoma, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación.

40 El casete de expresión que contiene los polinucleótidos proporcionados en el presente documento también puede comprender un gen marcador de selección para la selección de células transformadas. Los genes marcadores de selección se utilizan para la selección de células o tejidos transformados.

45 Cuando sea apropiado, las secuencias empleadas en los métodos y composiciones (es decir, el polinucleótido de interés, el agente de nucleasa, etc.) pueden optimizarse para aumentar la expresión en la célula. Es decir, los genes se pueden sintetizar usando codones preferidos en una célula de interés dada, incluidos, por ejemplo, codones preferidos de mamífero, codones preferidos de humano, codones preferidos de roedor, codones preferidos de ratón, codones preferidos de rata, etc., para expresión mejorada.

50 En una realización, el agente de nucleasa se expresa a partir de una construcción de expresión que comprende una secuencia de ácido nucleico que codifica una nucleasa y el ácido nucleico se une operativamente a un promotor activo en la célula.

VII. Métodos para elaborar animales no humanos genéticamente modificados (parte de la descripción que no describe la invención)

55 Se pueden generar animales no humanos genéticamente modificados empleando los diversos métodos divulgados en el presente documento. En algunos casos, el método para producir un animal no humano genéticamente modificado comprende: (1) modificar el genoma de una célula pluripotente no humana usando los métodos descritos en este documento; (2) seleccionar la célula pluripotente modificada genéticamente; (3) introducir la célula pluripotente modificada genéticamente en un embrión huésped no humano; y (4) implantar el embrión huésped que comprende la  
 60 célula pluripotente modificada genéticamente en una madre sustituta. Se genera una progenie de la célula pluripotente modificada genéticamente. La célula donante se puede introducir en un embrión huésped no humano en cualquier etapa, tal como la etapa de blastocisto o la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). Se genera descendencia que es capaz de transmitir la modificación genética a través de la línea germinal. La célula pluripotente puede ser una célula ES (p. ej., una célula ES de ratón o una célula ES de rata). Véase por ejemplo la  
 65 patente de Estados Unidos N° 7,294,754.

Las técnicas de transferencia nuclear también se pueden usar para generar animales mamíferos no humanos. Brevemente, los métodos para la transferencia nuclear pueden incluir las etapas de: (1) enucleación de un ovocito o proporcionar un ovocito enucleado; (2) aislar o proporcionar una célula o núcleo donante no humano para combinarlo con el ovocito no humano enucleado; (3) insertar la célula o el núcleo en el ovocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal no humano para formar un embrión; y (5) permitir que el embrión se desarrolle. En tales métodos, los ovocitos generalmente se recuperan de animales muertos, aunque también pueden aislarse de oviductos y/u ovarios de animales vivos. Los ovocitos se pueden madurar en una variedad de medios conocidos por los expertos en la materia antes de la enucleación. La enucleación del ovocito se puede realizar de varias maneras bien conocidas por los expertos en la materia. La inserción de la célula o el núcleo del donante en el ovocito enucleado para formar una célula reconstituida puede realizarse mediante microinyección de una célula del donante debajo de la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de CC a través del plano de contacto/fusión (electrofusión), por exposición de las células a sustancias químicas que promueven la fusión, tales como el polietilenglicol, o por medio de un virus inactivado, tal como el virus Sendai. Una célula reconstituida puede activarse por medios eléctricos y/o no eléctricos antes, durante y/o después de la fusión del ovocito nuclear donante y receptor. Los métodos de activación incluyen pulsos eléctricos, choques inducidos químicamente, penetración de espermatozoides, niveles crecientes de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (por medio de inhibidores de quinasa) en el ovocito. Las células reconstituidas activadas, o embriones no humanos, pueden cultivarse en un medio bien conocido por los expertos en la técnica y luego transferirse al útero de un animal. Véanse, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1 y la patente de los Estados Unidos N° 7,612,250.

Los métodos pueden comprender además métodos para producir un animal no humano de generación F0 que comprende: (1) identificar una célula ES no humana que comprende la modificación dirigida; (2) introducir la célula ES no humana que comprende la modificación dirigida en un embrión huésped no humano; y (3) gestar el embrión huésped no humano en una madre sustituta. La madre sustituta puede entonces producir el animal no humano de la generación F0 que comprende la modificación dirigida. El embrión huésped que comprende la célula pluripotente o totipotente genéticamente modificada (por ejemplo, una célula ES no humana) puede incubarse hasta la etapa de blastocisto y luego implantarse en una madre sustituta para producir un animal F0. Los animales que portan el locus genómico modificado genéticamente pueden identificarse mediante un ensayo de modificación del alelo (MOA) como se describe en el presente documento.

Los diversos métodos proporcionados en el presente documento permiten la generación de un animal F0 no humano modificado genéticamente en el que las células del animal F0 modificado genéticamente que comprende la modificación dirigida. Se reconoce que dependiendo del método usado para generar el animal F0, variará el número de células dentro del animal F0 que tienen la modificación genética dirigida. La introducción de las células ES del donante en un embrión en etapa de premórula de un organismo correspondiente (p. ej., un embrión de ratón en etapa de 8 células) mediante, por ejemplo, el método VELOCIMOUSE® permite un mayor porcentaje de la población celular del animal F0 para comprender células que tienen la modificación genética dirigida. En casos específicos, al menos 50 %, 60 %, 65 %, 70 %, 75 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100% de la contribución celular del animal F0 no humano comprende una población celular que tiene la modificación dirigida. En otros casos, al menos una o más de las células germinales del animal F0 tienen la modificación dirigida.

En algunos casos, las células del animal F0 modificado genéticamente son heterocigotas o heterocigotas compuestas para la modificación dirigida. Por ejemplo, las células del animal F0 modificado genéticamente pueden ser hemicigotas para la modificación dirigida. En otros casos, las células del animal F0 modificado genéticamente son homocigóticas para la modificación dirigida.

En algunos casos, el animal F0 generado por los métodos y composiciones divulgados en el presente documento puede cruzarse con un animal de tipo silvestre para generar una generación F1 que sea heterocigota para la modificación dirigida. A continuación, los animales de la generación F1 pueden cruzarse entre sí para generar un animal F2 homocigoto para la modificación dirigida. La progenie F1 se puede genotipar usando cebadores y/o sondas específicos para determinar si la modificación genética dirigida está presente.

#### VIII. Métodos de introducción de ácidos nucleicos y proteínas en las células

En el presente documento se proporcionan varios métodos y composiciones para permitir la introducción de un ácido nucleico en una célula. En algunos casos, el sistema empleado para introducir el ácido nucleico permite la integración dirigida en un locus genómico específico. Dichos sistemas emplean una variedad de componentes y, para facilitar la referencia, el término "sistema de integración genómica dirigida" incluye genéricamente todos los componentes necesarios para un evento de integración (por ejemplo, uno o más agentes de nucleasa, sitios de escisión de nucleasa, insertos de ácido nucleico, vectores de direccionamiento, loci genómicos diana y polinucleótidos de interés).

Los métodos proporcionados en el presente documento pueden comprender la introducción en una célula de uno o más polinucleótidos o construcciones polipeptídicas que comprenden uno o más componentes de un sistema de

integración genómica dirigida. "Introducir" incluye presentar a la célula la secuencia (polipéptido o polinucleótido) de tal manera que la secuencia obtenga acceso al interior de la célula. Los métodos proporcionados en el presente documento no dependen de un método particular para introducir un ácido nucleico o una proteína en la célula, solo que el ácido nucleico o la proteína accede al interior de al menos una célula. Los métodos para introducir ácidos nucleicos y proteínas en diversos tipos de células son conocidos en la técnica e incluyen, por ejemplo, métodos de transfección estables, métodos de transfección transitoria y métodos mediados por virus.

En algunos casos, las células empleadas en los métodos y composiciones tienen una construcción de ADN incorporada de forma estable en su genoma. "Incorporado de forma estable" o "introducido de forma estable" incluye la introducción de un polinucleótido en la célula de manera que la secuencia de nucleótidos se integre en el genoma de la célula y sea capaz de ser heredada por su descendencia. Puede usarse cualquier protocolo para la incorporación estable de las construcciones de ADN o los diversos componentes del sistema de integración genómica dirigida.

Los protocolos de transfección así como los protocolos para introducir polipéptidos o secuencias de polinucleótidos en las células pueden variar. Los métodos de transfección incluyen métodos de transfección basados en productos químicos que usan liposomas; nanopartículas; fosfato de calcio (Graham et al. (1973) *Virology* 52 (2): 456-67, Bacchetti et al. (1977) *Proc Natl Acad Sci USA* 74 (4): 1590-4, y Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. Nueva York: W. H. Freeman and Company, páginas 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilénimina. Los métodos no químicos incluyen la electroporación, la sonoporación y la transfección óptica. La transfección basada en partículas incluye el uso de una pistola de genes o transfección asistida por imán (Bertram (2006) *Current Pharmaceutical Biotechnology* 7, 277-28). También se pueden usar métodos virales para la transfección.

En algunos casos, la introducción de ácidos nucleicos o proteínas en una célula está mediada por electroporación, por inyección intracitoplasmática, por infección viral, por adenovirus, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos, o por Nucleofection<sup>MC</sup>.

La introducción de ácidos nucleicos o proteínas en la célula se puede realizar una o varias veces durante un período de tiempo. Por ejemplo, la introducción se puede realizar al menos dos veces durante un período de tiempo, al menos tres veces durante un período de tiempo, al menos cuatro veces durante un período de tiempo, al menos cinco veces durante un período de tiempo, al menos seis veces durante un período de tiempo, al menos siete veces durante un período de tiempo, al menos ocho veces durante un período de tiempo, al menos nueve veces durante un período de tiempo, al menos diez veces durante un período de tiempo, al menos once veces, al menos doce veces durante un período de tiempo, al menos trece veces durante un período de tiempo, al menos catorce veces durante un período de tiempo, al menos quince veces durante un período de tiempo, al menos dieciséis veces durante un período de tiempo, al menos diecisiete veces durante un período de tiempo, al menos dieciocho veces durante un período de tiempo, al menos diecinueve veces durante un período de tiempo, o al menos veinte veces durante un período de tiempo.

Cuando se introducen en la célula agentes de nucleasa y vectores de direccionamiento (p. ej., LTVEC), se pueden introducir simultáneamente. Alternativamente, el agente de nucleasa se puede introducir por separado de los vectores de direccionamiento. Por ejemplo, el agente de nucleasa puede introducirse antes de la introducción de los vectores de direccionamiento, o puede introducirse después de la introducción de los vectores de direccionamiento. Cuando se introducen dos o más LTVEC en la célula, se pueden introducir simultáneamente o, alternativamente, se pueden introducir por separado.

#### IX. Células y animales

Varias composiciones y métodos proporcionados en este documento emplean células, tales como células de un animal. Dichas células pueden ser células no humanas y pueden ser de un animal no humano. De acuerdo con los métodos de la invención, dichas células son células eucariotas, en particular, células de mamíferos. Una célula de mamífero puede ser, por ejemplo, una célula de mamífero no humano, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster, un fibroblasto o una célula CHO. La célula eucariota puede ser una célula totipotente, una célula pluripotente, tal como una célula pluripotente no humana (por ejemplo, una célula madre embrionaria de ratón (ES) o una célula ES de rata) o una célula pluripotente humana, o una célula no pluripotente. Las células totipotentes incluyen células indiferenciadas que pueden dar lugar a cualquier tipo celular, y las células pluripotentes incluyen células indiferenciadas que poseen la capacidad de convertirse en más de un tipo celular diferenciado. Tales células pluripotentes y/o totipotentes pueden ser, por ejemplo, células madre embrionarias (ES) o células similares a ES, tales como células madre pluripotentes inducidas (iPS). Las células madre embrionarias incluyen células totipotentes o pluripotentes derivadas de embriones no humanos que son capaces de contribuir a cualquier tejido del embrión en desarrollo al introducirse en un embrión no humano. Las células ES no humanas pueden derivarse de la masa celular interna de un blastocisto y son capaces de diferenciarse en células de cualquiera de las tres capas germinales de vertebrados (endodermo, ectodermo y mesodermo). Dichas células también pueden ser células madre hematopoyéticas o células madre neuronales.

Una célula eucariota también puede ser una célula que no es una célula somática primaria. Las células somáticas pueden incluir cualquier célula que no sea un gameto, una célula germinal, un gametocito o una célula madre indiferenciada.

5 Las células eucariotas también incluyen células primarias. Las células primarias incluyen células o cultivos de células que se han aislado directamente de un organismo, órgano o tejido. Las células primarias incluyen células que no se transforman ni son inmortales. Incluyen cualquier célula obtenida de un organismo, órgano o tejido que no haya pasado previamente por un cultivo de tejidos o que haya pasado previamente por un cultivo de tejidos pero que no sea capaz de pasar indefinidamente por un cultivo de tejidos. Dichas células pueden aislarse mediante técnicas convencionales e incluyen, por ejemplo, células somáticas, células hematopoyéticas, células endoteliales, células epiteliales, fibroblastos, células mesenquimatosas, queratinocitos, melanocitos, monocitos, células mononucleares, adipocitos, preadipocitos, neuronas, células gliales, hepatocitos, mioblastos esqueléticos y células de músculo liso. Por ejemplo, las células primarias pueden derivar de tejidos conectivos, tejidos musculares, tejidos del sistema nervioso o tejidos epiteliales.

15 Las células eucariotas también incluyen células inmortalizadas. Las células inmortalizadas incluyen células de un organismo multicelular que normalmente no proliferaría indefinidamente pero que, debido a una mutación o alteración, ha evadido la senescencia celular normal y, en cambio, puede continuar dividiéndose. Tales mutaciones o alteraciones pueden ocurrir naturalmente o ser inducidas intencionalmente. Los ejemplos de células inmortalizadas incluyen células de ovario de hámster chino (CHO), células de riñón embrionario humano (p. ej., células HEK 293) y células de fibroblastos embrionarios de ratón (p. ej., células 3T3). Numerosos tipos de células inmortalizadas son bien conocidas en la técnica.

20 Las células inmortalizadas o primarias incluyen células que normalmente se usan para cultivar o para expresar genes o proteínas recombinantes.

25 El término "animal", en referencia a células, células pluripotentes y/o totipotentes, células ES, células de donantes y/o embriones huésped, incluye mamíferos, peces y aves. Los mamíferos incluyen, por ejemplo, humanos, primates no humanos, monos, simios, gatos, perros, caballos, toros, ciervos, bisontes, ovejas, roedores (por ejemplo, ratones, ratas, hámsteres, conejillos de Indias), ganado (por ejemplo, especies bovinas tales como vacas, novillos, etc.; especies ovinas tales como ovejas, cabras, etc.; y especies porcinas tales como cerdos y jabalíes). Las aves incluyen, por ejemplo, pollos, pavos, avestruces, gansos, patos, etc. También se incluyen animales domésticos y animales de agricultura. El término "animal no humano" excluye a los humanos.

30 Las células pluripotentes y/o totipotentes de ratón pueden ser de una cepa 129, una cepa C57BL/6, una mezcla de 129 y C57BL/6, una cepa BALB/c o una cepa Swiss Webster. Los ejemplos de cepas 129 incluyen 129P1, 129P2, 129P3, 129X1, 129S1 (p. ej., 129S1/SV, 129SI/Svlm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6 (129/SvEvTac), 129S7, 129S8, 129T1 y 129T2. Véase, por ejemplo, Festing et al. (1999) Mammalian Genome 10:836). Los ejemplos de cepas C57BL incluyen C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Kal\_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr y C57BL/01a. Las células pluripotentes y/o totipotentes de ratón también pueden ser de una mezcla de una cepa 129 mencionada anteriormente y una cepa C57BL/6 mencionada anteriormente (p. ej., 50 % 129 y 50 % C57BL/6). Del mismo modo, las células pluripotentes y/o totipotentes de ratón pueden provenir de una mezcla de las cepas 129 mencionadas anteriormente o una mezcla de las cepas BL/6 mencionadas anteriormente (por ejemplo, la cepa 129S6 (129/SvEvTac)) Un ejemplo específico de una célula ES de ratón es una célula ES de ratón VGF1. Véase, por ejemplo, Auerbach et al. (2000) Biotechniques 29, 1024-1028, 1030, 1032.

35 Una célula pluripotente y/o totipotente de rata puede ser de cualquier cepa de rata, incluyendo, por ejemplo, una cepa de rata ACI, una cepa de rata Dark Agouti (DA), una cepa de rata Wistar, una cepa de rata LEA, una cepa de rata Sprague Dawley (SD) o una cepa de rata Fischer tal como Fisher F344 o Fisher F6. Las células pluripotentes y/o totipotentes de rata también se pueden obtener a partir de una cepa derivada de una mezcla de dos o más cepas mencionadas anteriormente. Por ejemplo, la célula pluripotente y/o totipotente de rata puede ser de una cepa DA o una cepa ACI. La cepa de ratas ACI se caracteriza por tener agutí negro, con el vientre y las patas blancas y un haplotipo *RT1<sup>av1</sup>*. Dichas cepas están disponibles en una variedad de fuentes, incluidas Harlan Laboratories. Un ejemplo de una línea de células ES de rata de una rata ACI es una célula ES de rata ACI.G1. La cepa de rata Dark Agouti (DA) se caracteriza por tener una capa de agouti y un haplotipo *RT1<sup>av1</sup>*. Dichas ratas están disponibles a través de una variedad de fuentes que incluyen Charles River y Harlan Laboratories. Los ejemplos de una línea de células ES de rata de una rata DA son la línea de células ES de rata DA.2B y la línea de células ES de rata DA.2C. En algunos casos, las células pluripotentes y/o totipotentes de rata proceden de una cepa de rata endogámica. Véase, por ejemplo, el documento U.S. 2014/0235933 A1, presentado el 20 de febrero de 2014, y el documento U.S. 2014/0310828 A1, presentado el 16 de abril de 2014, en su totalidad.

40 Los ejemplos de células pluripotentes humanas incluyen células ES humanas, células madre adultas humanas, células progenitoras humanas restringidas en el desarrollo y células madre pluripotentes inducidas humanas (iPS), tales como células iPS humanas cebadas y células iPS humanas sin modificar. Véase, por ejemplo, la solicitud de patente de los Estados Unidos N° 14/515,503, presentada el 15 de octubre de 2014. Las células madre pluripotentes inducidas

5 incluyen células madre pluripotentes que pueden derivarse directamente de una célula adulta diferenciada. Las células iPS humanas se pueden generar introduciendo conjuntos específicos de factores de reprogramación en una célula que pueden incluir, por ejemplo, Oct3/4, factores de transcripción de la familia Sox (p. ej., Sox1, Sox2, Sox3, Sox15), factores de transcripción de la familia Myc (p. ej., c-Myc, 1-Myc, n-Myc), factores de transcripción de la familia tipo Krüppel (KLF) (p. ej., KLF1, KLF2, KLF4, KLF5) y/o factores de transcripción relacionados, tales como NANOG, LIN28 y/o Glis1. Las células iPS humanas también se pueden generar, por ejemplo, mediante el uso de miARN, moléculas pequeñas que imitan las acciones de los factores de transcripción o especificadores de linaje. Las células iPS humanas se caracterizan por su capacidad para diferenciarse en cualquier célula de las tres capas germinales de vertebrados, por ejemplo, el endodermo, el ectodermo o el mesodermo. Las células iPS humanas también se caracterizan por su capacidad de propagarse indefinidamente en condiciones de cultivo *in vitro* adecuadas. Véase, por ejemplo, Takahashi y Yamanaka (2006) Cell 126: 663-676. Las células ES humanas cebadas y las células iPS humanas cebadas incluyen células que expresan características similares a las de las células epiblasticas posteriores a la implantación y están comprometidas con la especificación y la diferenciación del linaje. Las células ES humanas sin modificar y las células iPS humanas sin modificar incluyen células que expresan características similares a las de las células ES de la masa celular interna de un embrión previo a la implantación y no están comprometidas para la especificación de linaje. Véase, por ejemplo, Nichols y Smith (2009) Cell Stem Cell 4:487-492.

20 Las células que se han implantado en un embrión huésped no humano pueden denominarse "células donantes". La célula pluripotente y/o totipotente modificada genéticamente puede ser de la misma cepa que el embrión huésped o de una cepa diferente. Asimismo, la madre sustituta puede ser de la misma cepa que la célula pluripotente y/o totipotente modificada genéticamente y/o el embrión huésped, o la madre sustituta puede ser de una cepa diferente a la de la célula pluripotente y/o totipotente modificada genéticamente y/o el embrión huésped.

25 Se puede emplear una variedad de embriones huésped no humanos en los métodos y composiciones divulgados en el presente documento. Por ejemplo, las células pluripotentes y/o totipotentes que tienen la modificación genética dirigida pueden introducirse en un embrión no humano en etapa premórula (por ejemplo, un embrión en etapa de 8 células) de un organismo correspondiente. Véase, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008/0078000 A1. En otros casos, las células ES del donante pueden implantarse en un embrión huésped no humano en una etapa previa a la mórula, por ejemplo, etapa de 2 células, etapa de 4 células, etapa de 8 células, etapa de 16 células, etapa de 32 células o etapa de 64 células. El embrión huésped no humano también puede ser un blastocisto o puede ser un embrión de preblastocisto, un embrión en etapa de premórula, un embrión en etapa de mórula, un embrión en etapa de mórula no compactado o un embrión en etapa de mórula compactado. Cuando se emplea un embrión de ratón, la etapa del embrión huésped puede ser una Etapa 1 de Theiler (TS1), una TS2, una TS3, una TS4, una TS5 y una TS6, con referencia a las etapas de Theiler descritas en Theiler (1989) "The House Mouse: Atlas of Mouse Development", Springer-Verlag, Nueva York. Por ejemplo, la etapa de Theiler se puede seleccionar entre TS1, TS2, TS3 y TS4. En algunos casos, el embrión huésped no humano comprende una zona pelúcida y la célula donante es una célula ES que se introduce en el embrión huésped a través de un orificio en la zona pelúcida. En otros casos, el embrión huésped no humano es un embrión sin zona. En otros casos, se agrega el embrión huésped no humano en etapa de mórula.

40 X. Métodos de identificación de células con un locus genómico diana modificado

45 Algunos de los métodos anteriores comprenden además la identificación de una célula que tiene un locus genómico diana modificado (p. ej., un genoma modificado). Se pueden usar varios métodos para identificar las células que tienen una modificación dirigida, tal como una eliminación o una inserción. Dichos métodos pueden comprender la identificación de una célula que tiene la modificación dirigida en un locus diana. Se puede realizar un cribado para identificar dichas células con loci genómicos modificados.

50 La etapa de cribado puede comprender un ensayo cuantitativo para evaluar la modificación del alelo (MOA) de un cromosoma parental. Por ejemplo, el ensayo cuantitativo se puede realizar a través de una PCR cuantitativa, tal como una PCR en tiempo real (qPCR). La PCR en tiempo real puede utilizar un primer conjunto de cebadores que reconoce el locus diana y un segundo conjunto de cebadores que reconoce un locus de referencia no dirigido. El conjunto de cebadores puede comprender una sonda fluorescente que reconoce la secuencia amplificada.

55 En otros casos, las células que tienen la modificación genética dirigida se seleccionan utilizando métodos que incluyen, por ejemplo, análisis de transferencia Southern, secuenciación de ADN, análisis de PCR o análisis fenotípico. A continuación, dichas células se emplean en los diversos métodos y composiciones descritos en el presente documento.

60 Otros ejemplos de ensayos cuantitativos adecuados incluyen hibridación *in situ* mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación isotérmica de ADN, hibridación cuantitativa con una sonda o sondas inmovilizadas, Invader Probes®, ensayos MMP®, baliza molecular TAQMAN®, o tecnología de sonda Eclipse<sup>MC</sup> (vease, por ejemplo, el documento US2005/0144655).

65 La etapa de cribado generalmente también comprende ensayos específicos de brazo, que son ensayos utilizados para distinguir entre inserciones dirigidas correctas de un inserto de ácido nucleico en un locus genómico diana de

inserciones transgénicas aleatorias del inserto de ácido nucleico en ubicaciones genómicas fuera del locus genómico diana y también se utilizan para detectar el ensamblaje correcto de dos o más LTVEC superpuestos en una sola construcción. Los ensayos convencionales para la detección de modificaciones dirigidas, tales como la PCR de largo alcance o la transferencia Southern, vinculan el vector de direccionamiento insertado con el locus dirigido. Sin embargo, debido a sus grandes tamaños de brazo de homología, los LTVEC no permiten la detección mediante dichos ensayos convencionales. Para detectar el direccionamiento de LTVEC, se pueden usar ensayos de modificación de alelos (MOA) que incluyen ensayos de pérdida de alelos (LOA) y ganancia de alelos (GOA) (véase, por ejemplo, el documento US 2014/0178879 y Friendewey et al., (2010) *Methods Enzymol.* 476: 295-307). El ensayo de pérdida de alelo (LOA) invierte la lógica de cribado convencional y cuantifica el número de copias del locus nativo al que se dirigió la mutación. En un clon de células diana correctamente direccionado, el ensayo LOA detecta uno de los dos alelos nativos (para genes que no están en el cromosoma X o Y), siendo el otro alelo interrumpido por la modificación dirigida. El mismo principio se puede aplicar a la inversa como un ensayo de ganancia de alelo (GOA) para cuantificar el número de copias del vector de direccionamiento insertado. Por ejemplo, el uso combinado de los ensayos GOA y LOA revelará que un clon heterocigoto dirigido correctamente perdió una copia del gen diana nativo y ganó una copia del gen de resistencia a fármacos u otro marcador insertado.

Como ejemplo, la reacción en cadena de la polimerasa cuantitativa (qPCR) se puede usar como método de cuantificación de alelos, pero cualquier método que pueda distinguir de manera confiable la diferencia entre cero, una y dos copias del gen diana o entre cero, uno, y dos copias del inserto de ácido nucleico se puede usar para desarrollar un ensayo MOA. Por ejemplo, TAQMAN® se puede usar para cuantificar el número de copias de una plantilla de ADN en una muestra de ADN genómico, especialmente en comparación con un gen de referencia (véase, por ejemplo, el documento US 6,596,541). El gen de referencia se cuantifica en el mismo ADN genómico que el o los genes diana o el locus (loci). Por lo tanto, se realizan dos amplificaciones TAQMAN® (cada una con su respectiva sonda). Una sonda TAQMAN® determina el "Ct" (Ciclo de umbral) del gen de referencia, mientras que la otra sonda determina el Ct de la región del gen o locus (loci) direccionados que se reemplazan por una direccionamiento exitoso (es decir, un ensayo LOA). El Ct es una cantidad que refleja la cantidad de ADN de partida para cada una de las sondas TAQMAN®, es decir, una secuencia menos abundante requiere más ciclos de PCR para alcanzar el ciclo de umbral. Reducir a la mitad el número de copias de la secuencia molde para una reacción TAQMAN® dará como resultado un aumento de aproximadamente una unidad Ct. Las reacciones TAQMAN® en células en las que un alelo del gen o genes o locus (loci) ha sido reemplazado por recombinación homóloga darán como resultado un aumento de un Ct para la reacción TAQMAN® diana sin un aumento en el Ct para el gen de referencia en comparación con el ADN de células no diana. Para un ensayo GOA, se puede usar otra sonda TAQMAN® para determinar el Ct del inserto de ácido nucleico que está reemplazando el gen o genes o locus (loci) diana mediante direccionamiento exitoso.

Puede ser útil aumentar los ensayos LOA y GOA estándar para verificar el direccionamiento correcto por LTVEC. Por ejemplo, los ensayos LOA y GOA por sí solos pueden no distinguir los clones de células diana correctamente de los clones en los que una eliminación inducida por Cas del locus genómico diana coincide con la integración aleatoria de un LTVEC en otra parte del genoma. Debido a que la presión de selección en la célula diana se basa en el casete de selección, la integración transgénica aleatoria de LTVEC en otras partes del genoma generalmente incluirá el casete de selección y las regiones adyacentes de LTVEC, pero puede excluir regiones más distales de LTVEC. Por ejemplo, si una parte de un LTVEC se integra aleatoriamente en el genoma, y el LTVEC comprende un inserto de ácido nucleico de aproximadamente 5 kb o más de longitud con un casete de selección adyacente al brazo de homología 3', en algunos casos el brazo de homología 3', pero no el brazo de homología 5', se integrará transgénicamente con el casete de selección. Alternativamente, si el casete de selección adyacente al brazo de homología 5', en algunos casos el brazo de homología 5' pero no el brazo de homología 3' se integrará transgénicamente con el casete de selección. Por ejemplo, si los ensayos LOA y GOA se utilizan para evaluar la integración dirigida de LTVEC, y el ensayo GOA utiliza sondas contra el casete de selección o cualquier otra región única (sin brazo) de LTVEC, una eliminación heterocigota en el locus genómico diana combinado con una integración transgénica aleatoria de LTVEC dará la misma lectura que una integración dirigida heterocigota de LTVEC en el locus genómico diana. Para verificar el direccionamiento correcto por parte de LTVEC, se pueden utilizar ensayos específicos de brazo junto con ensayos LOA y/o GOA.

Los ensayos específicos de brazo determinan el número de copias de una plantilla de ADN en brazos de homología LTVEC. Dichos brazos de homología pueden incluir un brazo de homología de un LTVEC que no se superpone con otro LTVEC pero se corresponde con una secuencia diana en la célula (por ejemplo, el brazo de homología se superpone con la secuencia genómica dirigida en una célula de ratón (mArm)). Dichos brazos de homología también pueden incluir un brazo de homología superpuesto presente en dos LTVEC superpuestos (p. ej., secuencia humana superpuesta en el brazo de homología 3' de un primer LTVEC y el brazo de homología 5' de un segundo LTVEC (hArm)). Para los experimentos en los que se introducen múltiples LTVEC superpuestos en una célula, la detección generalmente comprende ensayos LOA, ensayos GOA para todas las secuencias insertadas únicas y ensayos específicos de brazo para todas las regiones de homología (es decir, entre LTVEC y la secuencia diana en la célula y entre dos diferentes LTVEC superpuestos). Como ejemplo, en el caso de tres LTVEC superpuestos introducidos en una célula de ratón para humanizar un locus diana de ratón de tipo silvestre, los números de copia esperados para la inserción dirigida heterocigótica serían los siguientes: 2 copias de mArm 5' (brazo de homología superpuesto con una secuencia diana de ratón 5'), 1 copia de hArm1 (secuencia superpuesta entre LTVEC 1 y 2), 1 copia de hArm2 (secuencia superpuesta entre LTVEC 2 y 3) y 2 copias de mArm 3' (brazo de homología superpuesto con secuencia

diana de ratón 3'). En el ejemplo anterior, los números de copias de mArm superiores a dos generalmente indicarían una integración de LTVEC transgénica aleatoriamente fuera del locus genómico diana en lugar de en el locus genómico diana, lo cual no es deseable. Los clones dirigidos correctamente retendrían un número de copias de mArm de dos. Además, los números de copias de mArm de menos de dos en dichos ensayos específicos de brazo generalmente indicaría grandes eliminaciones mediadas por Cas que se extienden más allá de la región diana para la eliminación, que también son indeseables. Del mismo modo, para las modificaciones dirigidas heterocigotas, los números de copia de 1 para hArm1 y hArm2 generalmente indicarían que los tres LTVEC se han ensamblado en una sola construcción.

Si diferentes versiones de una secuencia se asocian con un número de acceso en diferentes momentos, significa la versión asociada con el número de acceso en la fecha efectiva de presentación de esta solicitud. La fecha de presentación efectiva significa la anterior de la fecha de presentación real o la fecha de presentación de una solicitud prioritaria con referencia al número de acceso, si corresponde. Asimismo, si se publican diferentes versiones de una publicación, sitio web o similar en diferentes momentos, se entenderá la versión publicada más recientemente a la fecha efectiva de presentación de la solicitud, a menos que se indique lo contrario. Cualquier característica, etapa, elemento, realización o aspecto de la invención puede usarse en combinación con cualquier otro a menos que se indique específicamente lo contrario. Aunque la presente invención se ha descrito con cierto detalle a modo de ilustración y ejemplo con fines de claridad y comprensión, será evidente que se pueden practicar ciertos cambios y modificaciones.

## Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la materia una divulgación y una descripción completas de cómo fabricar y utilizar la presente invención. Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (p. ej., cantidades, temperatura, etc.), pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular promedio en peso, la temperatura está en grados centígrados y la presión es la atmosférica o cercana a ella.

Ejemplo 1: Direccionamiento del locus de TCR alfa con dos LTVEC en combinación con nucleasas con dedos de zinc

Se diseñó un sistema de direccionamiento doble de modo que dos vectores de direccionamiento grandes (LTVEC) modifican un locus genómico en una única etapa de direccionamiento. Como se representa en la Fig. 1, una célula que tiene una modificación heterocigótica del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina fue dirigida por el sistema de doble direccionamiento para generar células ES que comprenden segmentos de genes variables de I $\kappa$ K adicionales.

Este enfoque de doble direccionamiento, resumido en la Fig. 1, implica el direccionamiento doble o la coelectroporación de dos vectores de direccionamiento grande diferentes (LTVEC) junto con secuencias de nucleótidos que codifican una endonucleasa (p. ej., nucleasas con dedos de zinc) que crea una rotura de doble cadena en o cerca del locus diana en las células ES.

En este enfoque, el primer vector de direccionamiento grande (marcado como MAID 1710) comprendía un brazo de homología 3' de 30 kb que incluía la secuencia de los segmentos de los genes V $\kappa$ 1-5 y V $\kappa$ 1-6 humanos, una secuencia de 120 kb que comprendía los segmentos de los genes V $\kappa$ 3-7 a V $\kappa$ 3-15 humanos, y una región 5' de 20 kb ("región superpuesta") que comprendía el segmento del gen V $\kappa$ 1-16 humano. El segundo vector de direccionamiento grande (marcado como MAID 6600) comprendía una región de superposición 3' de 20 kb (región que comprende el segmento del gen V $\kappa$ 1-16 humano, igual que en el primer vector), una secuencia de 140 kb que comprende los segmentos de genes V $\kappa$ 1-17 a V $\kappa$ 2 -30 humanos, un casete de selección FRT-Ub-Neo-FRT y un brazo de homología de TCR A de ratón 3' de 15.5 kb.

Se diseñaron nucleasas con dedos de zinc (ZFN) que reconocen y escinden una secuencia diana dentro del gen de resistencia a higromicina para promover recombinación homóloga de los dos LTVEC en el locus de TCR A diana. Las células ES generadas en la Fig. 1 (MAID 6548, heterocigoto para todos los segmentos J $\kappa$  humanos y cuatro segmentos del gen V $\kappa$  humano funcional) se sometieron a electroporación con los dos vectores de direccionamiento grandes (MAID 6600 y MAID 1700 recortados) descritos anteriormente y dos plásmidos que expresan cada mitad del ZFN (1/2), que se unen a secuencias de reconocimiento en el gen de resistencia a la higromicina y catalizan una ruptura de doble cadena en el sitio diana (TGCGATCGCTGCGGCCGATcttagCCAGACGAGCGGGTTCGG (SEQ ID NO: 2); con el sitio de escisión en minúsculas) (véase la Tabla 1). Se diseñaron dos ZFN adicionales para dirigirse a la higromicina: ZFN (3/4) que se dirige al gen de la higromicina en la secuencia de nucleótidos CGCTGCGGCCGATCTtagcca GACGAGCCGGGTTTCGG (SEQ ID NO: 3); y ZFN (5/6) que se dirige al gen de la higromicina en la secuencia de nucleótidos AGCGTGTCGACCTGATGcagctcTCGGAGGGCGAAGAA (SEQ ID NO: 4) (véase la Tabla 1).

Tabla 1: Sitios de unión y escisión de la nucleasa con dedos de zinc de higromicina (no se muestra la cadena complementaria)

Nucleasa con	Mitad con	Secuencia de unión	Sitio de escisión
--------------	-----------	--------------------	-------------------



dedos de zinc	dedos de zinc	(5'-3')	(5'-3')
Hyg-ZFN(1,2)	Hyg-ZF1	TGCGATCGCTGCGGCCGA (SEQ ID NO: 5)	TCTTAG (SEQ ID NO: 11)
	Hyg-ZF2	CCGAACCCGCTCGTCTGG (SEQ ID NO: 6)	
Hyg-ZFN(3,4)	Hyg-ZF3	CGCTGCGGCCGATCT (SEQ ID NO: 7)	TAGCCA (SEQ ID NO: 12)
	Hyg-ZF4	CCGAACCCGCTCGTC (SEQ ID NO: 8)	
Hyg-ZFN(5,6)	Hyg-ZF5	AGCGTGTCCGACCTGATG (SEQ ID NO: 9)	CAGCTC (SEQ ID NO: 13)
	Hyg-ZF6	TTCTTCGCCCTCCGA (SEQ ID NO: 10)	

Los dos grandes vectores de direccionamiento se insertaron mediante recombinación homóloga en la secuencia de ADN reemplazando la región que contenía y rodeaba el casete de selección de Hyg. Las células ES resultantes contenían en el locus de TCR A endógeno un dominio variable de inmunoglobulina humana que comprende los segmentos de genes J $\kappa$ 1 a J $\kappa$ 5 y V $\kappa$ 4-1 a V $\kappa$ 2-30 humanos. La incorporación exitosa de los dos vectores de direccionamiento grandes se confirmó utilizando los ensayos TAQMAN® descritos anteriormente (Lie y Petropoulos, 1998. Curr. Opin Biotechnology 9:43-48.), usando sondas y cebadores indicados en la Fig. 1 y enumerados en la Tabla 2 a continuación (GOA = ganancia de alelo; LOA = pérdida de alelo; número de copias = verificar el número de copias de la secuencia para rastrear la integración transgénica frente a la integración dirigida; hArm1 = brazo de homología 3' de 30 kb del primer gran vector de direccionamiento (MAID 1710); hArm2 = superposición de 20 kb del primer (MAID 1710) y el segundo (MAID 6600) vectores de direccionamiento grandes, mArm = brazo de homología 5' de 15.5 kb del segundo vector de direccionamiento (MAID 6600), secuencias del control de ratón WT presentes en el locus de TCR A de ratón). Se utilizaron ensayos de PCR en tiempo real que reconocen secuencias en los brazos de homología de los LTVEC, denominados ensayos específicos de brazo, para verificar el direccionamiento correcto de LTVEC en el genoma del ratón. La determinación del número de copias de estos ensayos específicos de brazo brindó más aclaraciones para ayudar a distinguir los clones de ES dirigidos correctamente, que retienen, por ejemplo, un número de copias de mArm de dos, de los clones en los que una eliminación inducida por Cas9 del locus de ratón diana coincide con la integración aleatoria de los LTVEC en otras partes del genoma, en cuyo caso el número de copias de mArm sería tres (o más).

Tabla 2: Cebadores y sondas TAQMAN

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
HYG	LOA	TGCGGCCGATCTTA GCC (SEQ ID NO: 14)	ACGAGCGGGTTCG GCCCATTC (SEQ ID NO: 15)	TTGACCGATTTCCT GCGG (SEQ ID NO: 16)
HYG-U	LOA	CGACGTCTGTGCGA GAAGTTTCTG (SEQ ID NO: 17)	AGTTCGACAGCGT GTCCGACCTGA (SEQ ID NO: 18)	CACGCCCTCCTACA TCGAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGCGTACA CAAATCG (SEQ ID NO: 20)	CCGCTGGACCGAT GGCTGTGT (SEQ ID NO: 21)	GGGCGTCGGTTTCC ACTATC (SEQ ID NO: 22)
pgkP1(Hyg Promotor)	LOA	CAAATGGAAGTAG CACGTCTCACT (SEQ ID NO: 23)	CTCGTGCAGATGG ACAGCACCGC (SEQ ID NO: 24)	CCGCTGCCCAA GG (SEQ ID NO 25)
hlgK6	Número de copia de hArm1	GTCAAGCACTGCT GGCACAC (SEQ ID NO: 26)	AACCCTTGTGCTAT TGAATTGCTATGCT GTCAG (SEQ ID NO: 27)	TGTTGTAGACCCTC CGCCAC (SEQ ID NO: 28)
hlgK12 (inserto de MAID 1710)	GOA	TTGCCTTTCACA CCTGCAG (SEQ ID NO: 29)	CAGCCCATCCTGTC ACTTCGCTGGA (SEQ ID NO: 30)	TGGCCAAACAGTA CAGCTCAG (SEQ ID NO :31)
hlgK13	Número de copia de hArm2	TCAGTCAATCACCT TTCCAGC (SEQ ID NO: 32)	TCCCCAGGTAGCCT CATGAACCAATGTT (SEQ ID NO: 33)	CACATTACTGAGTC CCCACAGGG (SEQ ID NO: 34)
hlgK14	Número de copia de hArm2	CATTGTCAAAGAA GCACTGGAAATG (SEQ ID NO: 35)	ACCATTGCAGTTTA CCCACGGTTAGGA TTTTT (SEQ ID NO: 36)	TCTTGCAATGGGAT CATCAGATG (SEQ ID NO: 37)
Neo	GOA	GGTGGAGAGGCTA TTCGGC (SEQ ID NO: 38)	TGGGCACAACAGA CAATCGGCTG (SEQ ID NO: 39)	GAACACGGCGGCA TCAG (SEQ ID NO: 40)
hlgK15	GOA	CAGGTGCAAAGGT GACCACAG (SEQ ID NO: 41)	TGGGTCCTGCCCAT CCATGCA (SEQ ID NO: 42)	GGCAGCCTGAGTG TCAGAGC (SEQ ID NO: 43)
hlgK25	GOA	GTTCAGGCCCCAC AGACTCTC (SEQ ID NO: 44)	TCCTCTCTGGAGCA ACCATGAAGTCCC T (SEQ ID NO: 45)	CCTGAAGCCATGA GGGCG (SEQ ID NO: 46)
hUbc-D (promotor Neo)	GOA	AGGGTAGGCTCTC CTGAATCG (SEQ ID NO: 47)	ACAGGCGCCGGAC CTCTGGT (SEQ ID NO: 48)	CCAAAGAACTGA CGCCTCAC (SEQ ID NO: 49)

## ES 2 947 714 T3

Arm4 de TCR A	Número de copia de hArm	GCGCCACATGAAT TTGACCAG (SEQ ID NO: 50)	TGTACCCAATCTTC CAAAGAAAGAGCT G (SEQ ID NO: 51)	GGCATCCTGTCCTC CCTTC (SEQ ID NO: 52)
Parental 1540m1	Control de ratón WT	CAGTAAGGGAAGA GACTACAACAGCA T (SEQ ID NO: 53)	TGCACACTGCTCAC CACTGCAAGCTAT (SEQ ID NO: 54)	TGCTGGTGGCCCA TCT (SEQ ID NO: 55)
Parental 1540m3	Control de ratón WT	GAAGTCAGCTATG ATAGTGTGCAATGT A (SEQ ID NO: 56)	CAGCCCAGCAGCT GTGGGTTCTC (SEQ ID NO: 57)	GCTCAGGGAGAAC ACAGAACTTAGA (SEQ ID NO: 58)
hlgK5	Secuencia de MAID 6548	CCCCGTCCTCCTCC TTTTTC (SEQ ID NO: 59)	TCATGTCCATTAAC CCATTTACCTTTTG CCA (SEQ ID NO: 60)	TGCAAGTGCTGCCA GCAAG (SEQ ID NO: 61)

El locus dirigido resultante en células ES tenía las siguientes secuencias de unión, donde las secuencias de ratón están entre paréntesis, las secuencias humanas están en fuente normal, los sitios de clonación múltiples están en negrita y las secuencias Frt están en cursiva (Tabla 3).

5

Tabla 3: Secuencias de unión del locus resultante de direccionamiento de célula ES doble

Unión	SEQ ID NO	Secuencia
Tcra de ratón/Frt 5'	62	(GTCTTTTTTGTCTTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACACTCTTCA) <i>GAAGGGACTC</i> <b>CTGATTTCAAAGGGTACCGAAGTTCCTATT</b> <i>CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC</i>
Frt 3'/IgK humana	63	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGAACTTCCTAGGGTTTCACCG</i> <b>GTGGCGCGCTAACAGAGAGGAAAGTCAAA</b> TTATAAAGAATATGAGATTCAGAATTCTGA TTAACTGTGG
IgK humana/ Tcra de ratón	64	GATAAATTATTTTGTCTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAG <b>GCGGCC</b> <b>GCCACCGCGGTGGAGCTC</b> (AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

La modificación del cribado de la detección de alelos (MOA) de colonias de células ES aisladas dio como resultado la identificación de 27 clones dirigidos correctamente entre 960 colonias detectadas, para una eficacia de direccionamiento del 2.81%.

10

Una estrategia alternativa para generar loci de TCR A que comprenden segmentos de genes de región variable de inmunoglobulina adicionales implica el direccionamiento en serie con vectores de direccionamiento grandes sucesivos (véase, por ejemplo, la Fig. 2). Como tal, las células ES heterocigóticas para todos los segmentos del gen J $\kappa$  humano y cuatro segmentos del gen V $\kappa$  humano funcional (MAID 6548) fueron electroporados con un gran vector de direccionamiento que comprende, de 5' a 3': un brazo de homología de ratón 5' de 15.5 kb, un casete de selección Frt-Ub-Neo-Frt, un fragmento de 120 kb que comprende los segmentos de genes V $\kappa$ 3-7 a V $\kappa$ 3-15, y un brazo de homología humana 3' de 30 kb que comprende los segmentos de genes V $\kappa$ 1-5 y V $\kappa$ 1-6 (también presentes en la secuencia de MAID 6548). La incorporación exitosa se confirmó con los ensayos TAQMAN® descritos anteriormente, utilizando cebadores y sondas que se enumeran en la Tabla 2 anterior y se indican en la Fig. 2: Hyg, hlgK5, hlgK6, hlgK12, Neo, parental 1540m3, parental 1540m1. En particular, las sondas Arm4 y hlgK6 de TCRA se usaron como sondas específicas de brazo para validar el direccionamiento genómico correcto del LTVEC. También se puede utilizar un conjunto adicional de cebadores y sonda, hlgK10, para confirmar la incorporación exitosa: Cebador directo - CGATTATGACTGGTTAGGTAGAAAGGTG (SEQ ID NO: 65); Sonda - GCCACTGGTTTCTCCAAATGTTTTCAATCCAT (SEQ ID NO: 66); Cebador inverso - GGGAGTACTTGGAGATCCCTAAGC (SEQ ID NO: 67).

25

El locus diana resultante en células ES tenía las siguientes secuencias de unión, donde las secuencias de ratón están entre paréntesis, las secuencias humanas están en fuente normal, los sitios de clonación múltiples están en negrita y las secuencias de Frt están en cursiva (Tabla 4).

30

Tabla 4: Secuencia de unión de locus resultante del direccionamiento individual de células ES

Unión	SEQ ID NO	Secuencia
Tcra de ratón/Frt 5'	68	(TTGAGCTTCATCAAAGTCACATGGGTTAAA CTCTATGGAGTAGTCAGAACACACTCTTCA) <b>GAAGGGACTCCTGATTTCAAAGGGTACC</b> <i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAA</i> <i>AGTATAGGAACTTC</i>

Frnt 3'/IgK humana	69	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCCTAGGGTTTCACCG GTGGCGCGCCAGGACCCAGGCTCTGACACT CAGGCTGCCAATACAATTGCCATGAAGACA GATGTTGATG
IgK humana/ Tcra de ratón	64	GATAAATTATTTTGTGTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC GCCACCGCGGTGGAGCTC(AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

El cribado de MOA de colonias aisladas dio como resultado la identificación de 5 clones dirigidos correctamente entre 440 colonias cribadas (LTVEC solamente), para una eficacia de direccionamiento del 1.1%. Los resultados para el cribado de colonias aisladas dirigidas con LTVEC + ZFN o LTVEC + CRISPR-Cas9 se muestran en la Tabla 9.

5 Una vez completada el direccionamiento individual representado en la Fig. 2, las células ES pueden ser dirigidas sucesivamente con vectores de direccionamiento grandes que comprenden V<sub>k</sub> adicionales para sumar al repertorio completo de segmentos del gen V<sub>k</sub> de inmunoglobulina humana funcional.

10 En otras estrategias alternativas más, el direccionamiento doble o simple de segmentos de genes V<sub>k</sub> de Ig humana adicionales sucesivos puede lograrse usando esquemas de direccionamiento dobles (dos vectores de direccionamiento grandes) o sencillos (un vector de direccionamiento grande) que involucra la destrucción mediada por nucleasa con dedos de zinc o CRISPR de un casete o casetes de selección (por ejemplo, higromicina).

15 Las células ES diana descritas anteriormente se usan como células ES del donante y se introducen en un embrión en etapa de pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células, mediante el método VELOCIMOUSE® (véase, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008-0078000 A1). El embrión de ratón que comprende las células ES del donante se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES del donante. Los ratones F0  
20 completamente derivados de la célula ES del donante que portan de forma independiente un gen IgK V humano quimérico - Tcra C de ratón se identifican mediante genotipado utilizando un ensayo de modificación de alelo que detecta la presencia de secuencias genéticas únicas.

25 Ejemplo 2: Direccionamiento del gen de la higromicina con dos LTVEC en combinación con el sistema CRISPR/Cas

Los métodos de doble direccionamiento descritos en el Ejemplo 1 que utilizan nucleasas con dedos de zinc también se realizaron con un sistema CRISPR/Cas9.

30 Se diseñaron varios ARN guía (ARNg) para reconocer varias secuencias diana dentro del gen de resistencia a la higromicina (secuencia de reconocimiento CRISPR). Las secuencias de reconocimiento CRISPR dentro del gen de la higromicina fueron las siguientes: ARNg#1:

ACGAGCGGGTTCGGCCCCATTCGG (SEQ ID NO: 70); ARNg#6:

CTTAGCCAGACGAGCGGGTTCGG (SEQ ID NO: 71); ARNg#10:

35 GCCGATCTTAGCCAGACGAGCGGG (SEQ ID NO: 72); y ARNg#16:

CGACCTGATGCAGCTCTCGGAGG (SEQ ID NO: 73). Las ubicaciones de las secuencias de reconocimiento dentro del gen de la higromicina se representan en la Fig. 3, que representa la destrucción de la higromicina mediada por CRISPR/Cas al dirigirse al vector MAID 1545. Se cribaron ARNg#1, ARNg#6, ARNg#10 y ARNg#16 y se confirmó que se dirigían específicamente al gen de la higromicina (véase la Fig. 3). Los resultados del cribado primario utilizando los diversos ARNg específicos de higromicina se proporcionan en la Tabla 5.

Tabla 5. Resultados del cribado primario utilizando ARNg específicos de higromicina

ARNg/ZFN	#total de candidatos primarios (2 placas/electroporación)	Candidatos positivamente reconfirmados/total reconfirmados
ARNg1	5	2/2
ARNg6	6	1/1
ARNg10	19	5/5
ARNg16	91	8/8
ZFN 1/2	10	4/4

45 Las células ES, por ejemplo, la célula ES generada en la Fig. 1 (MAID 6548, heterocigotas para todos los segmentos J<sub>k</sub> humanos y cuatro segmentos del gen V<sub>k</sub> humano funcional) se sometieron a electroporación con dos grandes vectores de direccionamiento (descritos en el Ejemplo 1), junto con un solo vector o con múltiples vectores que codifican Cas9 y un ARNg (por ejemplo, ARNg#1, ARNg#6, ARNg#10 o ARNg#16), que reconocen y escinden un sitio diana dentro del gen de resistencia a la higromicina.

Se insertaron dos grandes vectores de direccionamiento mediante recombinación homóloga en la secuencia de ADN reemplazando la región que contenía y rodeaba el casete de selección de Hyg. La incorporación exitosa de los dos grandes vectores de direccionamiento se confirmó mediante ensayos TAQMAN®.

- 5 Las células ES diana descritas anteriormente se usarán como células ES del donante y se introducirán en un embrión en etapa de pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células, mediante el método VELOCIMOUSE® (véase, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008-0078000 A1). El embrión de ratón que comprende las células ES modificadas genéticamente se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES del donante.
- 10 Los ratones F0 completamente derivados de las células ES del donante se identificarán mediante genotipado usando un ensayo de modificación de alelo que detecta la presencia de secuencias genéticas únicas.

Ejemplo 3: Direccionamiento del locus de TCR alfa con tres LTVEC en combinación con nucleasas con dedos de zinc

- 15 Se diseñó un sistema de direccionamiento triple de modo que tres grandes vectores de direccionamiento (LTVEC) modifican un locus genómico en una sola etapa de direccionamiento. Como se representa en la Fig. 4, una célula que tiene una modificación heterocigótica del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina fue seleccionada como diana por el sistema de direccionamiento triple para generar células ES que comprenden segmentos de genes variables I $\kappa$ c adicionales.

- 20 Este enfoque de direccionamiento triple, resumido en la Fig. 4, implica el direccionamiento triple o la coelectroporación de tres vectores de direccionamiento grandes diferentes (LTVEC) (MAID 6647, MAID 6600 y MAID 1710) junto con secuencias de nucleótidos que codifican una endonucleasa (p. ej., nucleasas con dedos de zinc o Cas9 y ARNg) que crea una ruptura de doble cadena en o cerca del locus diana en las células ES.

- 25 En este enfoque, el primer vector de direccionamiento grande (marcado como MAID 1710) comprendía un brazo de homología 3' de 30 kb que incluye la secuencia de segmentos de genes V $\kappa$ 1-5 y V $\kappa$ 1-6 humanos, una secuencia de 120 kb que comprende segmentos de genes V $\kappa$ 3-7 a V $\kappa$ 3-15 humanos, y una región 5' de 20 kb ("región superpuesta") que comprendía el segmento del gen V $\kappa$ 1-16 humano. El segundo vector de direccionamiento grande (marcado como MAID 6600) comprendía una región de superposición 3' de 20 kb (región que comprende el segmento del gen V $\kappa$ 1-16 humano, igual que en el primer vector), una secuencia de 140 kb que comprende los segmentos de genes V $\kappa$ 1-17 a V $\kappa$ 2-24 humanos y una región 5' de 60 kb ("región superpuesta") que comprendía V $\kappa$ 3-25 a V $\kappa$ 2-30 humanos. El tercer vector de direccionamiento grande (marcado como MAID 6647) comprendía una región de superposición 3' de 60 kb (región que comprende V $\kappa$ 3-25 a V $\kappa$ 2-30 humanos, igual que en el segundo vector), una secuencia de 90 kb que comprende V $\kappa$ 3-31 a V $\kappa$ 2-40 humanos, un casete de selección FRT-Ub-Neo-FRT y un brazo de homología de TCR A de ratón 5' de 15.5 kb.

- 30 Se diseñaron nucleasas con dedos de zinc (ZFN) que reconocen y escinden una secuencia diana dentro del gen de resistencia a higromicina para promover recombinación homóloga de tres LTVEC en el locus de TCR A diana. Las células ES generadas en la Fig. 4 (MAID 6548, heterocigotas para todos los segmentos J $\kappa$  humanos y cuatro segmentos del gen VK humano funcional) se sometieron a electroporación con los tres vectores de direccionamiento grandes (MAID6600-recortados, MAID1700-recortados y MAID6647) descritos anteriormente y dos plásmidos que expresan cada mitad del ZFN (1/2), que se unen a secuencias de reconocimiento en el gen de resistencia a la higromicina y catalizan una rotura de doble cadena en el sitio diana (TGCGATCGCTGCGCCGAtcttag CCAGACGAGCGGGTTCGG (SEQ ID NO: 2); con el sitio de escisión en minúsculas) (véase la Tabla 1).

- 35 Los tres grandes vectores de direccionamiento se insertaron mediante recombinación homóloga en la secuencia de ADN reemplazando la región que contenía y rodeaba el casete de selección de Hyg. Las células ES resultantes contenían en el locus de TCR A endógeno un dominio variable de inmunoglobulina humana que comprende segmentos de genes humanos J $\kappa$ 1 a J $\kappa$ 5 y V $\kappa$ 4-1 a V $\kappa$ 2-40. La incorporación exitosa de los tres vectores de direccionamiento grandes se confirmó utilizando los ensayos TAQMAN® descritos anteriormente (Lie y Petropoulos, 1998. Curr. Opin. Biotechnology 9:43-48), usando sondas y cebadores indicados en la Fig. 4 y enumerados en la Tabla 2 anterior y en la Tabla 6 a continuación (GOA = ganancia de alelo; LOA = pérdida de alelo; número de copias = comprobar el número de copias de la secuencia para rastrear la integración transgénica frente a la integración dirigida; hArm1 = brazo de homología 3' de 30 kb del primer vector de direccionamiento grande (MAID 1710); hArm2 = superposición de 20 kb del primer vector de direccionamiento grande (MAID 1710) y el segundo (MAID 6600), hArm3 = superposición de 60 kb del segundo vector de direccionamiento (MAID 6600) y el tercero (MAID 6647), mArm = brazo de homología 5' de 15.5 kb del tercer vector de direccionamiento (MAID 6647), secuencias de control de ratón WT - presentadas en el locus de TCR A de ratón). Se utilizaron ensayos de PCR en tiempo real que reconocen secuencias en los brazos de homología de los LTVEC, denominados ensayos específicos de brazo, para verificar el direccionamiento correcto del LTVEC en el genoma de ratón. La determinación del número de copias de estos ensayos específicos de brazo brindó más aclaraciones para ayudar a distinguir los clones ES correctamente dirigidos, que retuvieron un número de copias de dos para la sonda de ratón (mArm) y un número de copias de uno para la sonda humana (hArm1), de clones en los que una eliminación inducida por Cas9 del locus diana de ratón coincide con la integración aleatoria de los LTVEC en otras partes del genoma, en cuyo caso habría un número de copias de tres (o más) para la sonda de ratón (mArm)

y un número copias de dos (o más) para la sonda humana (hArm1). Para detectar el ensamblaje correcto de los tres LTVEC mediante recombinación homóloga en el locus deseado, utilizamos ensayos TAQMAN® específicos de brazo. Los números de copias esperados, 1 para hArm2 y hArm3, indicaron que los tres LTVEC se ensamblaron en una sola construcción.

5

Tabla 6: Cebadores y sondas TAQMAN

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
HYG	LOA	TGCGGCCGATCTTA GCC (SEQ ID NO: 14)	ACGAGCGGGTTCG GCCCATTC (SEQ ID NO: 15)	TTGACCGATTCCTT GCGG (SEQ ID NO: 16)
HYG-U	LOA	CGACGTCTGTCGA GAAGTTTCTG (SEQ ID NO: 17)	AGTTCGACAGCGT GTCCGACCTGA (SEQ ID NO: 18)	CAGCCCTCCTACA TCGAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGCGTACA CAAATCG (SEQ ID NO: 20)	CCGTCTGGACCGAT GGCTGTGT (SEQ ID NO: 21)	GGCGTCGGTTTCC ACTATC (SEQ ID NO: 22)
hlgK6	Número de copias de hArm1	GTCAAGCACTGCT GGCACAC (SEQ ID NO: 26)	AACCCTTGTGCTAT TGAATTGCTATGCT GTCAG (SEQ ID NO: 27)	TGTTGTAGACCCTC CGCCAC (SEQ ID NO: 28)
hlgK12 (inserto de MAID 1710)	GOA	TTGCCTTTCTCACA CCTGCAG (SEQ ID NO: 29)	CAGCCCATCTGTAC TTCGCTGGA (SEQ ID NO:30)	TGGCCCAACAGTA CAGCTCAG (SEQ ID NO :31)
hlgK13	Número de copias de hArm2	TCAGTCAATCACCT TTCCCAGC (SEQ ID NO: 32)	TCCCCAGGTAGCCT CATGAACCAATGTT (SEQ ID NO: 33)	CACATTAAGTGTG CCCACAGGG (SEQ ID NO: 34)
hlgK14	Número de copias de hArm2	CATTGTCAAAGAA GCACTGGAAATG (SEQ ID NO: 35)	ACCATTGCAGTTTA CCCACGGTTAGGA TTTTT (SEQ ID NO: 36)	TCTTGAATGGGAT CATCAGATG (SEQ ID NO: 37)
Neo	GOA	GGTGGAGAGGCTA TTCGGC (SEQ ID NO: 38)	TGGGCACAACAGA CAATCGGCTG (SEQ ID NO: 39)	GAACACGGCGGCA TCAG (SEQ ID NO: 40)
hlgK15	GOA	CAGGTGCAAAGGT GACCACAG (SEQ ID NO: 41)	TGGGTCCTGCCAT CCATGCA (SEQ ID NO: 42)	GGCAGCCTGAGTG TCAGAGC (SEQ ID NO: 43)
hlgK25	Número de copias de hArm3	GTTCAGGCCCCAC AGACTCTC (SEQ ID NO: 44)	TCCTCTCTGGAGCA ACCATGAAGTTCCC T (SEQ ID NO: 45)	CCTGAAGCCATGA GGGCAG (SEQ ID NO: 46)
Arm4 de TCR A	Número de copia de hArm	GCGCCACATGAAT TTGACCAG (SEQ ID NO: 50)	TGTACCAATCTTC CAAAGAAAGAGCT G (SEQ ID NO: 51)	GGCATCTGTCTCTC CCTTC (SEQ ID NO: 52)
Parental 1540m1	Control de ratón WT	CAGTAAGGGAAGA GACTACAACAGCA T (SEQ ID NO: 53)	TGCACACTGCTCAC CACTGCAAGCTAT (SEQ ID NO: 54)	TGCTGGTGGCCCA TCT (SEQ ID NO: 55)
Parental 1540m3	Control de ratón WT	GAACTCAGCTATG ATAGTGTGCAATGT A (SEQ ID NO: 56)	CAGCCAGCAGCT GTGGTTCTC (SEQ ID NO: 57)	GCTCAGGGAGAAC ACAGAAGTCTAGA (SEQ ID NO: 58)
hlgK5	Secuencia de MAID 6548 (véase la Tabla 2)	CCCCGTCCTCTCC TTTTTC (SEQ ID NO: 59)	TCATGTCCATTAAC CCATTACCTTTTG CCCA (SEQ ID NO: 60)	TGCAAGTGTGCCA GCAAG (SEQ ID NO: 61)
hlgK22	Número de copias de hArm3	TGGCTCCAAGAACA GTTTGCC (SEQ ID NO: 74)	CCCTGACTTTGCTGCT CAACTCACAGCC (SEQ ID NO: 75)	GGTCCAGTGGAACT GCCATG (SEQ ID NO: 76)
hlgK21	GOA	CATTTGGCTACATAT CAAAGCCG (SEQ ID NO: 77)	CCTGAGCCAGGGAAC AGCCCACTGATA (SEQ ID NO: 78)	ACATGGCTGAGGCAG ACACC (SEQ ID NO: 79)
hlgK26	GOA	TGGGCGGTTATGCTA GTACCA (SEQ ID NO: 80)	TGGCTTACCCCTTTT GAAGGGCCC (SEQ ID NO: 81)	CACAGCTGAAGCAGG ATGAGC (SEQ ID NO: 82)
hlgK30	GOA	TCTCTGAGCAGCCAT CCCC (SEQ ID NO: 83)	TTCTCCTTGGTGTAG AGGGCACCAGC (SEQ ID NO: 84)	ACCAGGCATGGCAGA AAGG (SEQ ID NO: 85)

El locus diana resultante en células ES tenía las siguientes secuencias de unión, donde las secuencias de ratón están entre paréntesis, las secuencias humanas están en fuente normal, los sitios de clonación múltiples están en negrita y las secuencias de Frt están en cursiva (Tabla 7).

10

Tabla 7: Secuencias de unión de locus resultante del direccionamiento triple de células ES

Unión	SEQ ID NO	Secuencia
-------	-----------	-----------

## ES 2 947 714 T3

Tcra de ratón/Frt 5'	62	(GTCTTTTTGTTCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACACTCTTCA) <b>GAAGGGACTC</b> <b>CTGATTTCAAAGGGTACCGAAGTTCCTATT</b> <b>CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC</b> <b>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</b>
Frt 3'/IgK humana	86	<b>AAAGTATAGGAACTTCCTAGGGTTTCACCGGTG</b> <b>GCGCGCCTGAGTAGTGCTTTAGGTGTGTAATCA</b> <b>CCAAAGATTTAGTGAAGTCCCTGTGCAAGGAG</b> <b>GATAAATTATTTTGTGTCAGACAACAATAA</b> <b>AAATCAATAGCACGCCCTAAGAGCGGGC</b> <b>GCCACCGGGTGGAGCTC</b> (AGGTTTCCG GTTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)
IgK humana/ Tcra de ratón	64	

La modificación del cribado de alelos (MOA) de colonias de células ES aisladas dio como resultado una eficacia de direccionamiento del 0.4 % (véase la Tabla 8).

5            **Tabla 8. La modificación del cribado de alelos (MOA) da como resultado el direccionamiento con 3 LTVEC**

# de LTVEC	Nucleasa	Eliminar	Insertar	Eficiencia
3 LTVEC	ZFN	hyg	370 kb humana	0.4 %
3 LTVEC	ARNg#16/Cas9	hyg	370 kb humana	0.4 %
3 LTVEC	ninguna	hyg	370 kb humana	0 %

Las células ES diana descritas anteriormente se usan como células ES del donante y se introducen en un embrión en etapa de premórula, por ejemplo, un embrión de ratón en etapa de 8 células mediante el método VELOCIMOUSE® (véase, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008-0078000 A1). El embrión de ratón que comprende las células ES del donante se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES del donante. Los ratones F0 completamente derivados de la célula ES del donante que portan de forma independiente un gen IgK V humano quimérico-Tcra C de ratón se identificaron mediante genotipado utilizando un ensayo de modificación de alelo que detecta la presencia de secuencias de genes únicas.

15            Ejemplo 4: Direccionamiento del gen de la higromicina con tres LTVEC en combinación con el sistema CRISPR/Cas

Los métodos de direccionamiento triple descritos en el Ejemplo 3 que utilizan nucleasas con dedos de zinc también se realizaron con un sistema CRISPR/Cas9.

20            Se diseñaron varios ARN guía (ARNg) para reconocer varias secuencias diana dentro del gen de resistencia a la higromicina (secuencia de reconocimiento CRISPR). Las secuencias de reconocimiento CRISPR dentro del gen de la higromicina son las siguientes: ARNg#1:

25            ACGAGCGGGTTCGGCCCCATTCGG (SEQ ID NO: 70); ARNg#6:  
CTTAGCCAGACGAGCGGGTTCGG (SEQ ID NO: 71); ARNg#10:  
GCCGATCTTAGCCAGACGAGCGG (SEQ ID NO: 72); y ARNg#16:  
CGACCTGATGCAGCTCTCGGAGG (SEQ ID NO: 73). Las ubicaciones de las secuencias de reconocimiento dentro del gen de la higromicina se representan en la Fig. 3. Se cribaron ARNg#1, ARNg#6, ARNg#10 y ARNg#16 y se confirmó que se dirigían específicamente al gen de la higromicina (véanse la Fig. 3 y la Tabla 5).

35            Las células ES MAID 6548 (heterocigotas para todos los segmentos J $\kappa$  humano y cuatro segmentos del gen V $\kappa$  humano funcional) se sometieron a electroporación con tres grandes vectores de direccionamiento como se describe en el Ejemplo 3, junto con vectores que codifican Cas9 y ARNg#16, que reconocen y escinden un sitio diana dentro del gen de resistencia a la higromicina.

40            Se insertaron tres grandes vectores de direccionamiento mediante recombinación homóloga en la secuencia de ADN reemplazando la región que contenía y rodeaba el casete de selección de Hyg. La incorporación exitosa de los tres grandes vectores de direccionamiento se confirmó utilizando los ensayos TAQMAN® descritos en el Ejemplo 3.

El locus diana resultante en las células ES tenía las secuencias de unión que se muestran en la Tabla 7, donde las secuencias de ratón están entre paréntesis, las secuencias humanas están en fuente normal, los sitios de clonación múltiples están en negrita y las secuencias de Frt están en cursiva.

45            La modificación del cribado de alelos (MOA) de colonias de células ES aisladas dio como resultado una eficacia de direccionamiento del 0.4 % (véase la Tabla 8).

Las células ES diana descritas anteriormente se usarán como células ES del donante y se introducirán en un embrión en etapa de pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células, mediante el método VELOCIMOUSE®

(véase, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008-0078000 A1). El embrión de ratón que comprende las células ES modificadas genéticamente se incubaba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES del donante. Los ratones F0 completamente derivados de las células ES del donante se identificarán mediante el genotipado usando un ensayo de modificación del alelo que detecta la presencia de secuencias genéticas únicas.

Ejemplo 5: Mejora del direccionamiento de LTVEC a través de secuencias superpuestas entre dos LTVEC

Se empleó el sistema de doble direccionamiento descrito en el Ejemplo 1 para modificar un locus genómico en una sola etapa de direccionamiento utilizando dos vectores de direccionamiento grandes (LTVEC). Como se representa en la Fig. 1, una célula que tiene una modificación heterocigótica del locus de TCR alfa en el cromosoma 14 de ratón que comprende un casete de selección de higromicina fue dirigida por el sistema de doble direccionamiento para generar células ES que comprenden segmentos de genes variables de Igκ adicionales. Los dos LTVEC diferentes se coelectroporaron juntos en células madre embrionarias (ES) de ratón. Opcionalmente, se co-electroporó un ácido nucleico que codifica una endonucleasa (ya sea una nucleasa con dedos de zinc (ZFN) o CRISPR-Cas9) para crear una rotura de doble cadena en o cerca del locus diana.

Como en el Ejemplo 1, el LTVEC (marcado como MAID 1710) comprendía un brazo de homología 3' de 30 kb que incluía la secuencia de los segmentos de genes Vκ1-5 y Vκ1-6 humanos, una secuencia de 120 kb que comprendía los segmentos de genes Vκ3-7 a Vκ3-15, y una región 5' de 20 kb ("región superpuesta") que comprendía un segmento del gen Vκ1-16 humano. El segundo LTVEC (marcado como MAID 6600) comprendía una región de superposición 3' de 20 kb (región que comprende el segmento del gen Vκ1-16 humano, igual que en el primer vector), una secuencia de 140 kb que comprende los segmentos del gen Vκ1-17 a Vκ2-30 humano, un casete de selección FRT-Ub-Neo-FRT y un brazo de homología de TCR A de ratón 3' de 15.5 kb.

El direccionamiento exitoso dio como resultado la inserción de los dos LTVEC por recombinación homóloga en la secuencia de ADN reemplazando la región que contiene y rodea el casete de selección de Hyg. Las células ES resultantes contenían en el locus de TCR A endógeno un dominio variable de inmunoglobulina humana que comprende segmentos de genes Jκ1 a Jκ5 y Vκ4-1 a Vκ2-30 humanos. La incorporación exitosa de los dos vectores de direccionamiento grandes se confirmó utilizando los ensayos TAQMAN descritos anteriormente (Lie y Petropoulos, 1998. Curr. Opin. Biotechnology 9:43-48), utilizando las sondas y los cebadores indicados en la Fig. 1 y en la Tabla 2.

Como comparación, el sistema LTVEC único descrito en el Ejemplo 1 también se empleó para modificar el mismo locus genómico usando un LTVEC único, ya sea solo o en combinación con un ZFN o CRISPR-Cas9 (véase la Fig. 2). La incorporación exitosa fue confirmada por los ensayos TAQMAN descritos anteriormente, utilizando cebadores y sondas que se enumeran en la Tabla 2 anterior y se indican en la Fig. 2.

La Tabla 9 compara las eficiencias de direccionamiento en experimentos de direccionamiento usando el LTVEC único (solo, con ZFN o con Cas9), usando los dos LTVEC simultáneamente (solo, con ZFN o con Cas9), o usando los dos LTVEC más un tercer LTVEC simultáneamente (solo, con ZFN o con Cas9). Las eficiencias de direccionamiento que se presentan en la Tabla 9 son el porcentaje de clones ESC cribados que se determinó que estaban correctamente dirigidos a través del cribado inicial, el cribado de confirmación y el cribado de reconfirmación usando los cebadores y sondas TAQMAN de la Tabla 2. El direccionamiento con un solo LTVEC resultó en 1.1 % de clones dirigidos correctamente. La escisión con un ZFN aumentó la eficiencia de direccionamiento del LTVEC único al 4.4 %, y la escisión con CRISPR-Cas9 aumentó la eficiencia de direccionamiento del LTVEC único al 5.5 %. Sorprendentemente, el direccionamiento con 2 LTVEC que tenían 20 kb en secuencia superpuesta dio como resultado una eficacia de direccionamiento del 1.4 % incluso cuando no se utilizó nucleasa. La eficiencia de direccionamiento aumentó al 2.81% cuando se usó un ZFN y al 1.6% cuando se usó Cas9.

Tabla 9: Eficiencia de direccionamiento de LTVEC en mosaico coelectroporados

	MAID 6598: 2da inserción			MAID 6600: 2da y 3a inserciones			MAID 6647: 2da, 3a y 4a inserciones		
	LTVEC solo	LTVEC + ZFN	LTVEC + Cas9	2 LTVE solos	2 LTVEC + ZFN	2 LTVEC + Cas9	2 LTVEC solos	2 LTVEC + ZFN	2 LTVEC + Cas9
Eficiencia de direccionamiento	5 / 440 (1.1 %)	17 / 384 (4.4 %)	21 / 384 (5.5 %)	13 / 960 (1.4 %)	27 / 960 (2.8 %)	15 / 960 (1.6 %)	0 / 960 (0 %)	4 / 960 (0.4 %)	4 / 960 (0.4 %)

## REIVINDICACIONES

1. Un método para modificar un locus genómico diana en una célula de mamífero, que comprende:

- 5 (a) introducir en la célula de mamífero un agente de nucleasa o un polinucleótido que codifica al agente de nucleasa, en el que el agente de nucleasa produce una rotura de cadena sencilla o doble dentro del locus genómico diana;
- 10 (b) introducir en la célula de mamífero un primer vector de direccionamiento grande (LTVEC) que tiene al menos 10 kb de longitud y comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', y un segundo LTVEC que tiene al menos 10 kb de longitud y comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, y el primer brazo de homología 5' del primer LTVEC y el segundo brazo de homología 3' del segundo LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana, en el que la primera secuencia superpuesta es de al menos 1 kb, y en el que el locus genómico diana se modifica mediante la integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y
- 15 (c) seleccionar una célula de mamífero dirigida que comprenda el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrados en el locus genómico diana.

- 20 2. El método de la reivindicación 1, en el que el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico comprende ADN genómico, opcionalmente en el que el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico comprenden un alelo condicional, un polinucleótido que codifica un marcador de selección un gen informador, uno o más casetes de expresión, o un ácido nucleico flanqueado por secuencias diana de recombinación específicas del sitio, opcionalmente en el que el ADN genómico es homólogo u ortólogo a una secuencia dirigida a la eliminación en el locus genómico diana, y opcionalmente en el que la inserción del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico da como resultado el reemplazo de una secuencia de ácido nucleico no humano con una secuencia de ácido nucleico humano homólogo u ortólogo.

30 3. El método de la reivindicación 1 o 2, en el que:

- (I) el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son de una especie que es diferente de la especie de la célula del mamífero; y/o
- (II) el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son ácidos nucleicos humanos.

35 4. El método de la reivindicación 1 - 3, en el que:

- (I) el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de 50 kb a 500 kb; o
- 40 (II) la célula de mamífero direccionada comprende ADN genómico que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico juntos, que tienen un tamaño combinado que oscila de 5 kb a 500 kb.

5. El método de una cualquiera de las reivindicaciones 1 - 4, en el que:

- 45 (I) la primera secuencia superpuesta del primer LTVEC es idéntica a la primera secuencia superpuesta del segundo LTVEC; y/o
- (II) el tamaño de la primera secuencia superpuesta es de 1 kb a 70 kb; y/o
- (III) el tamaño de la primera secuencia superpuesta es de al menos 10 kb, opcionalmente en el que el tamaño de la primera secuencia superpuesta es de al menos 20 kb.

50 6. El método de una cualquiera de las reivindicaciones 1-5, en el que la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico diana da como resultado uno o más de:

- (I) una adición de una secuencia exógena en el locus genómico diana;
- 55 (II) una eliminación de una secuencia endógena en el locus genómico diana, opcionalmente en el que la eliminación es de 5 kb a 800 kb; y
- (III) una inserción, una eliminación, una mutación puntual, un intercambio de dominio, un intercambio de exón, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de los mismos.

60 7. El método de una cualquiera de las reivindicaciones 1 - 6, en el que:

- (I) la suma total de los brazos de homología 5' y 3' del primer LTVEC o el segundo LTVEC es de 10 kb a 150 kb; y/o
- 65 (II) el primer LTVEC es de al menos 50 kb y el segundo LTVEC es de al menos 50 kb, opcionalmente en el que el primer LTVEC es de al menos 100 kb y el segundo LTVEC es de al menos 100 kb.

8. El método para modificar un locus genómico diana en una célula de mamífero, que comprende:



(a) introducir en la célula de mamífero un agente de nucleasa o un polinucleótido que codifica al agente de nucleasa, en el que el agente de nucleasa produce una rotura de cadena sencilla o doble dentro del locus genómico diana;

(b) introducir en la célula de mamífero un primer vector de direccionamiento grande (LTVEC) que tiene al menos 10 kb de longitud y comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3', un segundo LTVEC que tiene una longitud de al menos 10 kb y comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y un tercer LTVEC que tiene una longitud de al menos 10 kb y comprende un tercer inserto de ácido nucleico flanqueado por un tercer brazo de homología 5' y un tercer brazo de homología 3', en el que el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia superpuesta homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia superpuesta homóloga al tercer brazo de homología 5' del tercer LTVEC, y el primer brazo de homología 5' del primer LTVEC y el tercer brazo de homología 3' del tercer LTVEC son homólogos a los segmentos genómicos correspondientes dentro del locus genómico diana, en el que la primera secuencia superpuesta tiene al menos 1 kb y la segunda secuencia superpuesta tienen al menos 1 kb, y

en el que el locus genómico diana se modifica mediante la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico entre los segmentos genómicos correspondientes; y

(c) seleccionar una célula de mamífero dirigida que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrados en el locus genómico diana.

9. El método de la reivindicación 8, en el que el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico y/o el tercer inserto de ácido nucleico comprenden ADN genómico, opcionalmente en el que el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico y/o el tercer inserto de ácido nucleico comprenden un alelo condicional, un polinucleótido que codifica un marcador de selección, un gen informador, uno o más casetes de expresión, o un ácido nucleico flanqueado por secuencias diana de recombinación específicas del sitio, opcionalmente en el que el ADN genómico es homólogo u ortólogo a una secuencia que se dirige para eliminación en el locus genómico diana, y opcionalmente en el que la inserción del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico da como resultado el reemplazo de una secuencia de ácido nucleico no humana con una secuencia de ácido nucleico humana homóloga u ortóloga.

10. El método de la reivindicación 8 o 9, en el que:

(I) uno o más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico son de una especie que es diferente de la especie de la célula de mamífero; y/o

(II) uno o más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico son ácidos nucleicos humanos.

11. El método de una cualquiera de las reivindicaciones 8 - 10, en el que:

(I) el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico es de 50 kb a 700 kb; o

(II) la célula de mamífero dirigida comprende ADN genómico que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico juntos, que tienen un tamaño combinado que oscila de 5 kb a 700 kb.

12. El método de una cualquiera de las reivindicaciones 8 - 11, en el que:

(I) la primera secuencia superpuesta del primer LTVEC es idéntica a la primera secuencia superpuesta del segundo LTVEC, y/o la segunda secuencia superpuesta del segundo LTVEC es idéntica a la segunda secuencia superpuesta del tercer LTVEC; y/o

(II) el tamaño de la primera secuencia superpuesta es de 1 kb a 70 kb, y/o el tamaño de la segunda secuencia superpuesta es de 1 kb a 70 kb; y/o

(III) el tamaño de la primera secuencia superpuesta es de al menos 10 kb, y/o el tamaño de la segunda secuencia superpuesta es de al menos 10 kb, opcionalmente en el que el tamaño de la primera secuencia superpuesta es de al menos 20 kb y/o el tamaño de la segunda secuencia superpuesta es de al menos 20 kb.

13. El método de una cualquiera de las reivindicaciones 8-12, en el que la integración de uno o más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico en el locus genómico diana da como resultado uno o más de:

(I) una adición de una secuencia exógena en el locus genómico diana;

(II) una eliminación de una secuencia endógena en el locus genómico diana, opcionalmente en el que la eliminación es de 5 kb a 800 kb; o

(III) una inserción, una eliminación, una mutación puntual, un intercambio de dominio, un intercambio de exón, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de los mismos.

14. El método de una cualquiera de las reivindicaciones 8 - 13, en el que:

- 5 (I) la suma total de los brazos de homología 5' y 3' del primer LTVEC, el segundo LTVEC, o el tercer LTVEC es de 10 kb a 150 kb; y/o  
(II) el primer LTVEC es de al menos 50 kb, el segundo LTVEC es de al menos 50 kb, y el tercer LTVEC es de al menos 50 kb, opcionalmente en el que el primer LTVEC es de al menos 100 kb, el segundo LTVEC es de al menos 100 kb, y el tercer LTVEC es de al menos 100 kb.

10 15. El método de una cualquiera de las reivindicaciones 1-14, en el que el agente de nucleasa es una nucleasa con dedos de zinc (ZFN), una nucleada efectora similar al activador de transcripción (TALEN), o una meganucleasa.

15 16. El método de una cualquiera de las reivindicaciones 1-15, en el que el agente de nucleasa es una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) y un ARN guía (ARNg), opcionalmente en el que la proteína Cas es Cas9.

17. El método de una cualquiera de las reivindicaciones 1 - 16, en el que la célula de mamífero es:

- 20 (I) una célula humana;  
(II) una célula no humana;  
(III) una célula de roedor, opcionalmente en el que el roedor es un ratón o una rata;  
(IV) una célula pluripotente, opcionalmente en el que la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS);  
25 (V) una célula madre hematopoyética;  
(VI) una célula madre neuronal; o  
(VII) una célula de fibroblasto.

30 18. El método de la reivindicación 17, en el que la célula de mamífero es una célula ES de ratón o una célula ES de rata, opcionalmente que comprende además introducir la célula ES de ratón o la célula ES de rata en un embrión huésped no humano.

FIG. 1

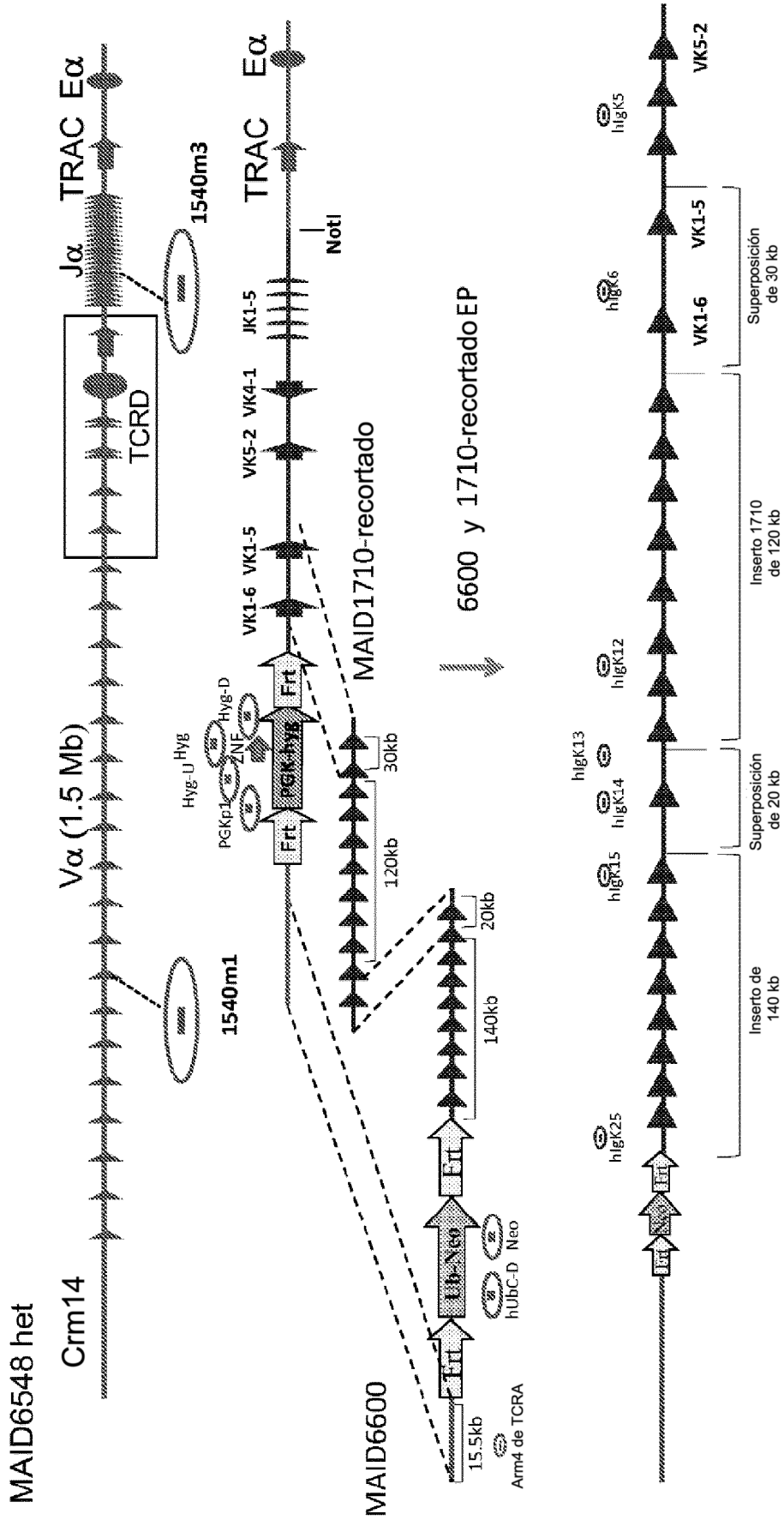


FIG. 2

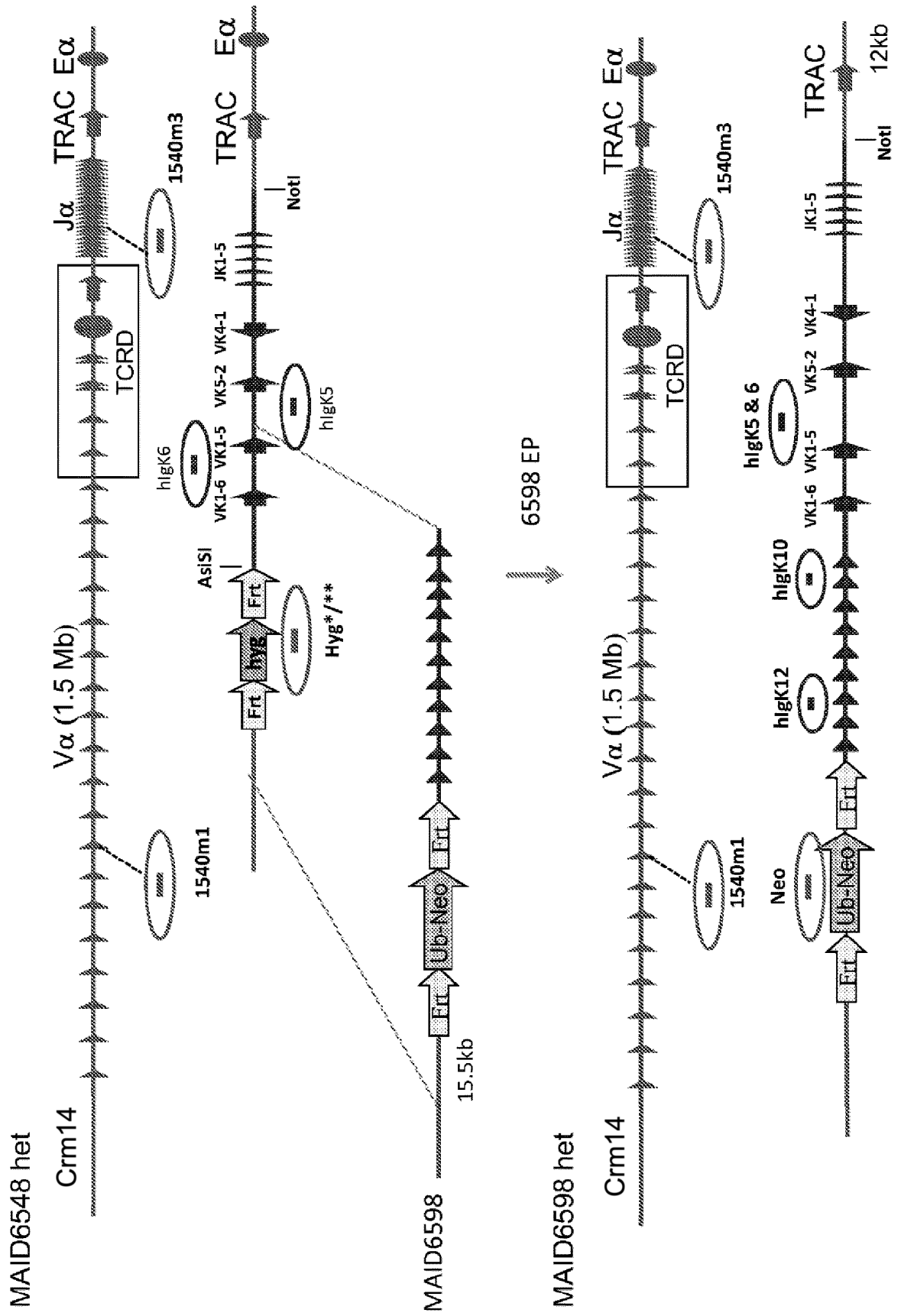
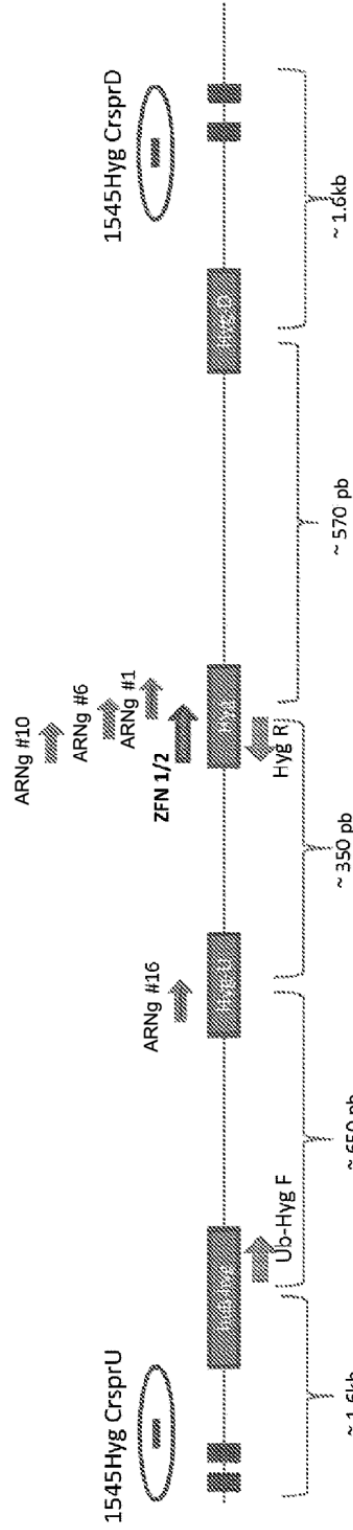


FIG. 3

Destrucción por CRISPR de higromicina en MAID 1545



ARNg/Znf	# Total de Candidatos Primarios (2 placas/EP)	Candidatos positivamente reconfirmados/total reconfirmados
ARNg1	5	2/2
ARNg6	6	1/1
ARNg10	19	5/5
ARNg16	91	8/8
Znf 1/2	10	4/4

FIG. 4

