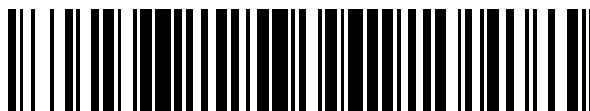


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 760 508**

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

86 Fecha de presentación y número de la solicitud internacional: **18.12.2015 PCT/US2015/066681**

87 Fecha y número de publicación internacional: **23.06.2016 WO16100819**

96 Fecha de presentación y número de la solicitud europea: **18.12.2015 E 15828604 (7)**

97 Fecha y número de publicación de la concesión europea: **09.10.2019 EP 3232774**

54 Título: **Métodos y composiciones para la modificación genética dirigida a través de la transformación múltiple en una sola etapa**

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:

**14.05.2020**

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 J.**

74 Agente/Representante:

**SÁEZ MAESO, Ana**

ES 2 760 508 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 la transformación múltiple en una sola etapa

5 Antecedentes de la invención

La recombinación homóloga con el uso de vectores de transformación diseñados para añadir, delecionar, 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 la técnica de la modificación del genoma a través de la recombinación homóloga ha avanzado considerablemente en las últimas dos décadas, aún persisten las dificultades para lograr una frecuencia de transformación aceptable con el uso de vectores de transformación muy grandes, los LTVEC, en muchas circunstancias, por ejemplo, cuando una gran parte de un genoma de roedor se reemplaza con un gran fragmento genómico humano, o cuando se dirige a ciertos tipos de células, por ejemplo, fibroblastos u otras células somáticas. Jallepalli y otros, "Securin is required for chromosomal stability in human cells the extremely early embryonic lethality of mice homozygously deleted for MAD2 and BUB3, preventing the evaluation of chromosome loss rates in proliferating somatic" Cell, 2001, páginas 445-457 describen un método para mejorar la recombinación homóloga en un locus genómico objetivo en una célula. El documento US20140309487 describe un método para modificar un locus genómico objetivo en una célula con el uso de un LTVEC. Kujipers y otros, "One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in Saccharomyces cerevisiae", Fems Yeast Research, vol. 13, núm. 8, 2019, páginas 769-781 describen un método para modificar un locus genómico objetivo en una célula de levadura.

15

20

Resumen

25

Se proporcionan métodos y composiciones para modificar un locus genómico objetivo dentro de una célula a través de un sistema de transformación que utiliza dos o más vectores de transformación que son capaces de recombinarse entre sí para formar un único segmento contiguo de ácido nucleico. Los vectores de transformación son vectores de transformación grandes (LTVEC). Los LTVEC son cada uno de al menos 10 kb de tamaño.

30

La invención es como se define en las reivindicaciones.

En un primer aspecto de la invención, se proporciona allí un método para modificar un locus genómico objetivo en una célula de mamífero, que comprende:

35

(a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro del locus genómico objetivo;

(b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de 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 es de 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',

40

en donde el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de al menos 100 kb,

en donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición 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 objetivo,

45

en donde la primera secuencia de superposición es de al menos 1 kb, y

en donde el locus genómico objetivo se modifica mediante la integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula transformada que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo.

50

En otro aspecto de la invención, se proporciona un método para modificar un locus genómico objetivo en una célula de mamífero, que comprende:

55

(a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro del locus genómico objetivo;

(b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de 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 es de 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 es de 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',

60

en donde el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico es de al menos 100 kb,

65

en donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia de superposición 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 objetivo,

en donde la primera secuencia de superposición es de al menos 1 kb y la segunda secuencia de superposición es de al menos 1 kb, y

en donde el locus genómico objetivo 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 transformada que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo.

La invención proporciona métodos para modificar un locus genómico objetivo en una célula, que comprende: (a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo; (b) introducir en la célula un primer vector de transformación 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 donde 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 objetivo y el primer brazo de homología 3' del primer LTVEC y el segundo brazo 5' del segundo LTVEC son homólogos entre sí o respectivamente a otros brazos de homología 5' y 3' de uno o más LTVEC adicionales, cada uno que comprende un inserto de ácido nucleico adicional flanqueado por otro brazo de homología 5' y otro brazo de homología 3', en donde el locus genómico objetivo se modifica mediante la integración del primer inserto de ácido nucleico, uno o más insertos de ácido nucleico adicionales de uno o más LTVEC adicionales si están presentes, y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula transformada que comprende el primer inserto de ácido nucleico, uno o más insertos de ácido nucleico adicionales si están presentes, y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo. El primer LTVEC, el segundo LTVEC, y uno o más LTVEC adicionales son cada uno de al menos 10 kb de tamaño. En algunos de tales métodos, los LTVEC adicionales son uno o más de otros LTVEC que, cuando están presentes, se insertan entre el primer LTVEC y el segundo LTVEC.

La invención proporciona además métodos de transformación doble para modificar un locus genómico objetivo en una célula, que comprende (a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro del locus genómico objetivo; (b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de al menos 10 kb de tamaño 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 es de 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 donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición 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 objetivo, en donde el locus genómico objetivo se modifica mediante integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula transformada que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo.

Opcionalmente, el primer inserto de ácido 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 objetivo.

En algunos de tales métodos, la célula es una célula humana. En otros de tales métodos, la célula es una célula no humana. En algunos de tales métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). En algunos de tales 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 los métodos anteriores, el agente tipo nucleasa es una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa. En algunos de los métodos anteriores, el agente tipo nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) y un ARN guía (ARNg). Opcionalmente, la proteína Cas es Cas9.

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.

En algunos métodos de la descripción, 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 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, 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 de la descripción, la célula transformada 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 varía de aproximadamente 5 kb a aproximadamente 500 kb.

En algunos métodos, la primera secuencia de superposición del primer LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 % o 99,9 % idéntica a la primera secuencia de superposición del segundo LTVEC. En algunos métodos, el tamaño de la primera secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En algunos métodos, el tamaño de la primera secuencia de superposición es 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 objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. Opcionalmente, la delección de la secuencia endógena en el locus genómico objetivo 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 la transformación en comparación con el uso de un solo LTVEC. Opcionalmente, el aumento en la eficiencia de la transformación 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 el 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 o el 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. 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, 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, de aproximadamente 120 kb a aproximadamente 150 kb.

La invención proporciona además métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo por cualquiera de los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta, en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

La invención proporciona además métodos de transformación triple para modificar un locus genómico objetivo en una célula, que comprende: (a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble

cadena dentro del locus genómico objetivo; (b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de 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 es de 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 es de 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 donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia de superposición 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 objetivo, en donde el locus genómico objetivo 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 objetivo que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo.

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 núcleo 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 objetivo.

En algunos de tales métodos, la célula es una célula humana. En otros de tales métodos, la célula es una célula no humana. En algunos de tales métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). En algunos de tales 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 tales métodos, el agente tipo nucleasa es una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa. En algunos de tales métodos, el agente tipo nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (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 tales métodos de la descripción, 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 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 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 tales métodos de la descripción, la célula transformada 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 varían de aproximadamente 5 kb a aproximadamente 700 kb.

En algunos de tales métodos, la primera secuencia de superposición del primer LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntica a la primera secuencia de superposición del segundo LTVEC, y/o la segunda secuencia de superposición del segundo LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 % o 99,9 % idéntica a la segunda secuencia de superposición de la tercera LTVEC. En algunos de tales métodos, el tamaño de la primera secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb, y/o el tamaño de la segunda secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En algunos de tales métodos, el tamaño de la primera secuencia de superposición es de al

menos 10 kb o al menos 20 kb, y/o el tamaño de la segunda secuencia de superposición es de al menos 10 kb o al menos 20 kb.

5 En algunos de tales 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 objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o  
10 una combinación de estos. Opcionalmente, la delección de la secuencia endógena en el locus genómico objetivo 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 de 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  
15 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 de la descripción, 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 de la descripción, 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 tales métodos de la descripción, 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 de la descripción, 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.

La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo por cualquiera de los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta; en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

La descripción proporciona además métodos para mejorar la recombinación homóloga en un locus genómico objetivo en una célula, que comprende introducir en la célula un primer ácido nucleico y un segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de nucleótidos superpuesta. En algunos de tales métodos, la recombinación homóloga se mejora en comparación con los métodos en los que se introduce un solo ácido nucleico en la célula.

55 En algunos de tales métodos, la recombinación homóloga se mejora en el locus genómico objetivo sin usar un agente tipo nucleasa. Algunos de tales métodos comprenden además introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena en o cerca del locus genómico objetivo. En algunos de tales métodos, el agente tipo nucleasa es una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa. En algunos de tales métodos, el agente tipo nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) y un ARN guía (ARNg). Opcionalmente, la proteína Cas es Cas9.

65 En algunos de tales métodos, el método mejora la recombinación homóloga del primer ácido nucleico, el segundo ácido nucleico, o ambos en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el primer ácido nucleico se introduce sin el segundo ácido nucleico. Algunos de tales métodos mejoran la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el segundo ácido nucleico se

introduce sin el primer ácido nucleico. Opcionalmente, la mejora de la recombinación homóloga es 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.

5 En algunos de tales métodos, la secuencia de superposición del primer ácido nucleico es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntica a la secuencia de superposición del segundo ácido nucleico. En algunos de tales métodos, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. Opcionalmente, la secuencia de superposición es 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, o de aproximadamente 65 kb a aproximadamente 70 kb. En algunos de tales métodos, la secuencia de superposición es de 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, o al menos 70 kb. Opcionalmente, la secuencia de superposición es de al menos 20 kb.

20 En algunos de tales métodos, el primer ácido nucleico es un vector de transformación 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 no comprende una secuencia de nucleótidos que es homóloga al locus genómico objetivo excepto la secuencia de superposición. Opcionalmente, el primer vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb, de aproximadamente 2 kb a aproximadamente 5 kb, o de aproximadamente 5 kb a aproximadamente 10 kb. Opcionalmente, el primer vector de transformación es un primer vector de transformación grande (LTVEC). Opcionalmente, el primer LTVEC es de al menos 10 kb de longitud. Opcionalmente, el primer vector de transformación es un primer vector de transformación grande (LTVEC) que varía de aproximadamente 20 kb a aproximadamente 200 kb. Opcionalmente, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb.

30 En algunos de tales métodos, el primer ácido nucleico es un primer vector de transformación 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 transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'. Opcionalmente, el primer vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb, de aproximadamente 2 kb a aproximadamente 5 kb, o de aproximadamente 5 kb a aproximadamente 10 kb, y/o el segundo vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb, de aproximadamente 2 kb a aproximadamente 5 kb, o de aproximadamente 5 kb a aproximadamente 10 kb. Opcionalmente, el primer vector de transformación es un primer vector de transformación grande (LTVEC) y/o el segundo vector de transformación es un segundo vector de transformación grande (LTVEC). Opcionalmente, el primer LTVEC es de al menos 10 kb de longitud y/o el segundo LTVEC es de al menos 10 kb de longitud. Opcionalmente, el primer vector de transformación es un primer vector de transformación grande (LTVEC) que varía de aproximadamente 20 kb a aproximadamente 200 kb, y/o el segundo vector de transformación es un segundo vector de transformación grande (LTVEC) que varía de aproximadamente 20 kb a aproximadamente 200 kb. Opcionalmente, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb, y/o el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. Opcionalmente, la suma total de los brazos de homología 5' y 3' del primer LTVEC o el segundo LTVEC es de 10 kb a aproximadamente 200 kb. Opcionalmente, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb.

55 En algunos métodos, la secuencia de superposición se localiza en el extremo 3' del primer ácido nucleico y el extremo 5' de la segunda secuencia de ácido nucleico. En algunos métodos, la secuencia de nucleótidos superpuesta facilita el reclutamiento de la maquinaria de recombinación al locus genómico objetivo.

60 En algunos de tales métodos, la célula es una célula humana. En otros de tales métodos, la célula es una célula no humana. En algunos de tales métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). En algunos de tales 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.

La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo por cualquiera de los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta; en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

Se proporcionan métodos y composiciones para modificar un locus genómico objetivo dentro de una célula a través de un sistema de transformación que utiliza dos o más vectores de transformación que son capaces de recombinarse entre sí para formar un único segmento contiguo de ácido nucleico. En diversas modalidades, los vectores de transformación son vectores de transformación grandes (LTVEC). Los LTVEC son cada uno de al menos 10 kb de tamaño.

En un ejemplo, se proporciona un método para modificar un locus genómico objetivo en una célula. Un método de ese tipo comprende introducir en una célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo, introducir un primer vector de transformación 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 donde el primer brazo de homología 5' del primer LTVEC y el segundo homología 3' el brazo del segundo LTVEC es homólogo a los segmentos correspondientes dentro del locus objetivo y el primer brazo de homología 3' del primer LTVEC y el segundo brazo 5' del segundo LTVEC son homólogos entre sí o respectivamente a otros 5' y 3' homología brazos de uno o más LTVEC adicionales, que comprenden cada uno un inserto adicional flanqueado por un brazo de homología 5' adicional y un brazo de homología 3' adicional; en donde el locus genómico objetivo se modifica mediante la integración del primer inserto, uno o más insertos adicionales de uno o más LTVEC adicionales si están presentes, y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes. El primer LTVEC, el segundo LTVEC, y uno o más LTVEC adicionales son cada uno de al menos 10 kb de tamaño. El método comprende además seleccionar una célula transformada que comprende el primer inserto de ácido nucleico, uno o más insertos de ácido nucleico adicionales si están presentes, y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo. En tales métodos, los LTVEC adicionales son uno o más de otros LTVEC que, cuando están presentes, se insertan entre el primer LTVEC y el segundo LTVEC.

En otro ejemplo, se proporciona un método de transformación doble para modificar un locus genómico objetivo en una célula. Un método de ese tipo comprende introducir en una célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo, introducir un primer vector de transformación 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'. El primer LTVEC es de al menos 10 kb de tamaño y el segundo LTVEC es de al menos 10 kb de tamaño. En un método de ese tipo, el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición 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 correspondientes dentro del locus objetivo, de modo que el locus genómico objetivo 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 transformada que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo.

En algunos de tales 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 objetivo.

En otra modalidad, se proporciona un método de transformación triple para modificar un locus genómico objetivo en una célula. Un método de ese tipo comprende introducir en una célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo, introducir un primer vector de transformación 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'. El primer LTVEC es de al menos 10 kb de tamaño, el segundo LTVEC tiene al menos 10 kb de tamaño, y el tercer LTVEC tiene al menos 10 kb de tamaño. En un método de ese tipo, el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia de superposición 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 correspondientes dentro del locus objetivo, de manera que el locus genómico objetivo se modifica mediante la integración del primer, segundo, y tercer insertos de ácido nucleico entre los segmentos genómicos correspondientes. El método comprende además seleccionar una célula transformada que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo.



5 En algunos de tales 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, 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 objetivo.

10 En una modalidad, la célula es una célula pluripotente. En otra modalidad, la célula pluripotente es una célula madre embrionaria (ES). En algunas modalidades, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En otra modalidad, la célula es una célula madre pluripotente inducida (iPS).

15 En una modalidad, el locus genómico objetivo está en el genoma de la célula. En otra modalidad, el locus genómico objetivo está en el ADN extracromosómico dentro de la célula.

20 En una modalidad, la célula es una célula de fibroblasto.

25 En algunos métodos, la célula es una célula no humana. En otros métodos, la célula es de un humano. En algunas modalidades la célula es una célula de mamífero. En otra modalidad, la célula de mamífero es de un roedor. En algunos casos, el roedor es un ratón, una rata, o un hámster.

30 En algunos de los métodos anteriores, el agente tipo nucleasa se expresa a partir de un constructo de expresión que comprende una secuencia de ácido nucleico que codifica una nucleasa, y en donde el ácido nucleico se enlaza operativamente a un promotor activo en la célula. En otros métodos, el agente tipo nucleasa se expresa a partir de un ARNm que codifica la nucleasa. En algunos de tales métodos, la nucleasa es una nucleasa de dedos de zinc (ZFN). En otros de tales métodos, la nucleasa es una nucleasa efectora de tipo activador de la transcripción (TALEN). Aun en otros métodos, la nucleasa es una meganucleasa.

35 En algunos de los métodos anteriores, el agente tipo nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) y un ARN guía (ARNg). En algunos de tales métodos, la proteína Cas es Cas9.

40 En algunos de los métodos anteriores, 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 una modalidad, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y/o el tercer inserto de ácido nucleico son de una especie diferente. En algunos métodos, uno de más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico son de una especie 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. En otra modalidad, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y/o el tercer inserto de ácido nucleico son ácidos nucleicos humanos. En algunos métodos, uno de más del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico son ácidos nucleicos humanos.

45 En un ejemplo, el tamaño combinado del primer 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 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, o de aproximadamente 450 kb a aproximadamente 500 kb. En otra modalidad, el tamaño combinado del primer y el segundo inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 500 kb. Aun en otra modalidad, el tamaño combinado del primer y el segundo inserto de ácido nucleico es de aproximadamente 300 kb.

50 En algunas modalidades, la célula transformada comprende un ADN genómico que comprende el primer y el segundo inserto de ácido nucleico juntos que varían de aproximadamente 5 kb a aproximadamente 500 kb.

55 En un ejemplo, el tamaño combinado del primer, el segundo, y el tercer insertos 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 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 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.

5 En algunos ejemplos, la célula transformada comprende un ADN genómico que comprende el primer, el segundo, y el tercer insertos de ácido nucleico juntos que varían de aproximadamente 5 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. En algunas modalidades, el tamaño combinado del primer, el segundo, y el tercer insertos de ácido nucleico es de aproximadamente 400 kb.

10 En algunos de los métodos anteriores, la primera secuencia de superposición del primer LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntica a la primera secuencia de superposición del segundo LTVEC. En algunos de los métodos anteriores, la segunda secuencia de superposición del segundo LTVEC es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntica a la segunda secuencia de superposición del tercer LTVEC. En cualquiera de los métodos anteriores, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En una modalidad específica, la secuencia de superposición es de al menos 10 kb. En otra modalidad específica, la secuencia de superposición es de al menos 20 kb. En algunos de los métodos anteriores, la primera secuencia de superposición y/o la segunda secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En algunos métodos, la primera secuencia de superposición y/o la segunda secuencia de superposición es al menos 10 kb o al menos 20 kb.

20 En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. En algunos métodos, la integración de uno o más del primer, el segundo y el tercer insertos de ácido nucleico en el locus genómico objetivo resulta uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos.

25 En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado una adición de una secuencia exógena en el locus genómico objetivo. En una modalidad, la integración del primer, el segundo y/o el tercer insertos de ácido nucleico en el locus genómico objetivo da como resultado una adición de una secuencia exógena en el locus genómico objetivo.

30 En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado una delección de una secuencia endógena en el locus genómico objetivo. En otra modalidad, la integración del primer, el segundo, y/o el tercer insertos de ácido nucleico en el locus genómico objetivo da como resultado una delección de una secuencia endógena en el locus genómico objetivo. En algunos de tales métodos la delección de la secuencia endógena en el locus genómico objetivo 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, 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.

35 En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos insertos en el locus genómico objetivo da como resultado una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. Aun en otra modalidad, la integración del primer, el segundo, y/o el tercer insertos de ácido nucleico en el locus genómico objetivo da como resultado una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos.

40 En algunos de los métodos anteriores, el uso combinado del primer LTVEC y el segundo LTVEC da como resultado una mayor eficiencia de transformación en comparación con el uso de un solo LTVEC. Opcionalmente, el aumento en la eficiencia de la transformación 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.

45 En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC o el 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 algunas modalidades, 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 otras modalidades, la suma total de los brazos de homología 5' y 3' del primer LTVEC, el segundo LTVEC, o el tercer LTVEC es de aproximadamente 1 kb a aproximadamente 5 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 120 kb, o de aproximadamente 120 kb a 150 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de aproximadamente 1 kb a aproximadamente 5 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 120 kb, o de aproximadamente 120 kb a 150 kb; la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de aproximadamente 1 kb a aproximadamente 5 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 120 kb, o de aproximadamente 120 kb a 150 kb; y la suma total de los brazos de homología 5' y 3' del tercer LTVEC es de aproximadamente 1 kb a aproximadamente 5 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 120 kb, o de aproximadamente 120 kb a 150 kb.

Además, en los ejemplos se proporciona un método para producir un animal no humano de generación F0. Un método de ese tipo comprende introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo por cualquiera de los métodos anteriores, y gestar el embrión huésped no humano en una madre sustituta de manera que la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

La descripción proporciona además métodos para mejorar la recombinación homóloga en un locus genómico objetivo en una célula, que comprende introducir en la célula un primer ácido nucleico y un segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de nucleótidos superpuesta. En algunos de tales métodos, la recombinación homóloga se mejora en comparación con los métodos en los que se introduce un solo ácido nucleico en la célula. En algunos de tales métodos, la recombinación homóloga se mejora en el locus genómico objetivo sin usar un agente tipo nucleasa. Otros de tales métodos comprenden además introducir en la célula un agente tipo nucleasa que produce una ruptura de la cadena monocatenaria o bicatenaria en o cerca del locus genómico objetivo.

En un aspecto de la descripción, se proporciona un método para mejorar la recombinación homóloga en un locus genómico en una célula sin usar un agente tipo nucleasa, que comprende introducir en la célula un primer ácido nucleico y un segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de nucleótidos superpuesta.

En un ejemplo, el método mejora la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el primer ácido nucleico se introduce sin el segundo ácido nucleico. En un ejemplo, el método mejora la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el segundo ácido nucleico se introduce sin el primer ácido nucleico. En un ejemplo, el método aumenta la recombinación homóloga del primer y el segundo ácido nucleico en el locus genómico objetivo.

En un ejemplo, la mejora de la recombinación homóloga es al menos 0,5 veces, 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 una modalidad, la secuencia de superposición del primer ácido nucleico es homóloga a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntico a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es 100 % idéntica a esa secuencia de superposición del segundo ácido nucleico.

En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En algunos métodos, la secuencia de superposición es de al menos 20 kb. En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 5 kb. En una modalidad, la secuencia de superposición es de aproximadamente 5 kb a aproximadamente 10 kb. En una modalidad, la secuencia de superposición es de aproximadamente 10 kb a aproximadamente 15 kb. En una modalidad, la secuencia de superposición es de aproximadamente 15 kb a aproximadamente 20 kb. En una modalidad, la secuencia de superposición es de

aproximadamente 20 kb a aproximadamente 25 kb. En una modalidad, la secuencia de superposición es de aproximadamente 25 kb a aproximadamente 30 kb. En una modalidad, la secuencia de superposición es de aproximadamente 30 kb a aproximadamente 35 kb. En una modalidad, la secuencia de superposición es de aproximadamente 35 kb a aproximadamente 40 kb. En una modalidad, la secuencia de superposición es de aproximadamente 40 kb a aproximadamente 45 kb. En una modalidad, la secuencia de superposición es de aproximadamente 45 kb a aproximadamente 50 kb. En una modalidad, la secuencia de superposición es de aproximadamente 50 kb a aproximadamente 55 kb. En una modalidad, la secuencia de superposición es de aproximadamente 55 kb a aproximadamente 60 kb. En una modalidad, la secuencia de superposición es de aproximadamente 60 kb a aproximadamente 65 kb. En una modalidad, la secuencia de superposición es de aproximadamente 65 kb a aproximadamente 70 kb.

En una modalidad, la secuencia de superposición es de al menos 5 kb. En una modalidad, la secuencia de superposición es de al menos 10 kb. En una modalidad, la secuencia de superposición es de al menos 15 kb. En una modalidad, la secuencia de superposición es de al menos 20 kb. En una modalidad, la secuencia de superposición es de al menos 25 kb. En una modalidad, la secuencia de superposición es de al menos 30 kb. En una modalidad, la secuencia de superposición es de al menos 35 kb. En una modalidad, la secuencia de superposición es de al menos 40 kb. En una modalidad, la secuencia de superposición es de al menos 45 kb. En una modalidad, la secuencia de superposición es de al menos 50 kb. En una modalidad, la secuencia de superposición es de al menos 55 kb. En una modalidad, la secuencia de superposición es de al menos 60 kb. En una modalidad, la secuencia de superposición es de al menos 65 kb. En una modalidad, la secuencia de superposición es de al menos 70 kb.

En una modalidad, el primer ácido nucleico es un vector de transformación 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 no comprende una secuencia de nucleótidos que sea homóloga al locus genómico excepto la secuencia de superposición.

En una modalidad, el segundo ácido nucleico es un segundo vector de transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y el primer ácido nucleico no comprende una secuencia de nucleótidos que sea homóloga al locus genómico excepto la secuencia de superposición.

En una modalidad, el primer ácido nucleico es un primer vector de transformación 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 transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'. En una modalidad, el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico son fragmentos superpuestos de un ácido nucleico contiguo.

En un ejemplo, el vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb. En un ejemplo, el vector de transformación es de aproximadamente 2 kb a aproximadamente 5 kb. En un ejemplo, el vector de transformación es de aproximadamente 5 kb a aproximadamente 10 kb.

El vector de transformación es un vector de transformación grande (LTVEC). En algunos métodos, el vector de transformación es un LTVEC que varía de aproximadamente 20 kb a aproximadamente 200 kb. En algunos métodos, el primer vector de transformación es un primer LTVEC que varía de aproximadamente 20 kb a aproximadamente 200 kb, y/o el segundo vector de transformación es un segundo LTVEC que varía de aproximadamente 20 kb a aproximadamente 200 kb. En una modalidad, el LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. En algunos métodos, el primer vector de transformación es un primer LTVEC y/o el segundo vector de transformación es un segundo LTVEC. En algunos métodos, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. En algunos métodos, el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. En algunos métodos, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb, y el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb.

5 En una modalidad, la suma total de los brazos de homología 5' y 3' del LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb.

10 En una modalidad, la secuencia de superposición se localiza en el extremo 3' del primer ácido nucleico y el extremo 5' de la segunda secuencia de ácido nucleico. En un ejemplo, la secuencia de superposición se localiza en el extremo 5' de la primera secuencia de ácido nucleico y el extremo 3' de la segunda secuencia de ácido nucleico.

15 En una modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son de una especie diferente. En otra modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son ácidos nucleicos humanos. En algunos métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son de una especie 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.

20 En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. En algunos métodos, la integración de uno o más del primer, el segundo y el tercer insertos de ácido nucleico en el locus genómico objetivo resulta uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos.

25 En una modalidad, la integración del primer y/o segundo insertos en el locus genómico da como resultado una adición de una secuencia exógena en el locus genómico.

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

35 En otra modalidad, la integración del primer y/o segundo insertos de ácido nucleico en el locus genómico da como resultado una delección de una secuencia endógena en el locus genómico objetivo. En una modalidad, la delección de la secuencia endógena en el locus genómico objetivo 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, 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.

40 En algunos métodos, la célula es una célula humana. En otros métodos, la célula es una célula no humana. En algunos métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS), en donde el método no implica el uso de un embrión humano para propósitos industriales o comerciales. En algunos 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.

45 En una modalidad, la célula es una célula pluripotente. En otra modalidad, la célula pluripotente es una célula madre embrionaria (ES). En algunas modalidades, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En otra modalidad, la célula es una célula madre pluripotente inducida (iPS).

50 En una modalidad, el locus genómico objetivo está en el genoma de la célula. En otra modalidad, el locus genómico objetivo está en el ADN extracromosómico dentro de la célula.

55 En una modalidad, la célula es una célula de fibroblasto.

60 En algunos métodos, la célula es una célula no humana. En otros métodos, la célula es de un humano. En algunas modalidades la célula es una célula de mamífero. En otra modalidad, la célula de mamífero es de un roedor. En algunos casos, el roedor es un ratón, una rata, o un hámster.

En algunos de los métodos anteriores, la secuencia de nucleótidos superpuesta facilita el reclutamiento de la maquinaria de recombinación al locus genómico objetivo.

5 La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo mediante los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta, en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

10 En otro aspecto de la descripción, se proporciona un método para mejorar la recombinación homóloga en un locus genómico objetivo en una célula con un agente tipo nucleasa, que comprende introducir en la célula: (i) un primer ácido nucleico y un segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de nucleótidos superpuesta; y (ii) un agente tipo nucleasa que produce una ruptura de simple o doble cadena en o cerca del locus genómico.

15 En un ejemplo, el método mejora la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el primer ácido nucleico se introduce sin el segundo ácido nucleico. En un ejemplo, el método mejora la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el segundo ácido nucleico se introduce sin el primer ácido nucleico. En un ejemplo, el método aumenta la recombinación homóloga del primer y el segundo ácido nucleico en el locus genómico objetivo. En un ejemplo, la mejora de la recombinación homóloga es al menos 0,5 veces, 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.

20 En una modalidad, la secuencia de superposición del primer ácido nucleico es homóloga a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntica a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es 100 % idéntica a la secuencia de superposición del segundo ácido nucleico.

35 En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 5 kb. En una modalidad, la secuencia de superposición es de aproximadamente 5 kb a aproximadamente 10 kb. En una modalidad, la secuencia de superposición es de aproximadamente 10 kb a aproximadamente 15 kb. En una modalidad, la secuencia de superposición es de aproximadamente 15 kb a aproximadamente 20 kb. En una modalidad, la secuencia de superposición es de aproximadamente 20 kb a aproximadamente 25 kb. En una modalidad, la secuencia de superposición es de aproximadamente 25 kb a aproximadamente 30 kb. En una modalidad, la secuencia de superposición es de aproximadamente 30 kb a aproximadamente 35 kb. En una modalidad, la secuencia de superposición es de aproximadamente 35 kb a aproximadamente 40 kb. En una modalidad, la secuencia de superposición es de aproximadamente 40 kb a aproximadamente 45 kb. En una modalidad, la secuencia de superposición es de aproximadamente 45 kb a aproximadamente 50 kb. En una modalidad, la secuencia de superposición es de aproximadamente 50 kb a aproximadamente 55 kb. En una modalidad, la secuencia de superposición es de aproximadamente 55 kb a aproximadamente 60 kb. En una modalidad, la secuencia de superposición es de aproximadamente 60 kb a aproximadamente 65 kb. En una modalidad, la secuencia de superposición es de aproximadamente 65 kb a aproximadamente 70 kb.

40 En una modalidad, la secuencia de superposición es de al menos 5 kb. En una modalidad, la secuencia de superposición es de al menos 10 kb. En una modalidad, la secuencia de superposición es de al menos 15 kb. En una modalidad, la secuencia de superposición es de al menos 20 kb. En una modalidad, la secuencia de superposición es de al menos 25 kb. En una modalidad, la secuencia de superposición es de al menos 30 kb. En una modalidad, la secuencia de superposición es de al menos 35 kb. En una modalidad, la secuencia de superposición es de al menos 40 kb. En una modalidad, la secuencia de superposición es de al menos 45 kb. En una modalidad, la secuencia de superposición es de al menos 50 kb. En una modalidad, la secuencia de superposición es de al menos 55 kb. En una modalidad, la secuencia de superposición es de al menos 60 kb. En una modalidad, la secuencia de superposición es de al menos 65 kb. En una modalidad, la secuencia de superposición es de al menos 70 kb.

45 En una modalidad, el primer ácido nucleico es un vector de transformación 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 no comprende una secuencia de nucleótidos que sea homóloga al locus genómico excepto la secuencia de superposición.

50 En una modalidad, el segundo ácido nucleico es un segundo vector de transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y el primer ácido nucleico no comprende una secuencia de nucleótidos que sea homóloga al locus genómico objetivo excepto la secuencia de superposición.

En una modalidad, el primer ácido nucleico es un primer vector de transformación 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 transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'. En una modalidad, el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico son fragmentos superpuestos de un ácido nucleico contiguo.

En un ejemplo, el vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb. En un ejemplo, el vector de transformación es de aproximadamente 2 kb a aproximadamente 5 kb. En un ejemplo, el vector de transformación es de aproximadamente 5 kb a aproximadamente 10 kb.

El vector de transformación es un vector de transformación grande (LTVEC). En algunos métodos, el vector de transformación es un vector de transformación grande que varía de aproximadamente 10 kb a aproximadamente 200 kb. En algunos métodos, el primer vector de transformación es un primer LTVEC y/o el segundo vector de transformación es un segundo LTVEC. En algunos métodos, el primer vector de transformación es un primer vector de transformación grande que varía de aproximadamente 10 kb a aproximadamente 200 kb, y/o el segundo vector de transformación es un segundo vector de transformación grande que varía de aproximadamente 10 kb a aproximadamente 200 kb. En una modalidad, el LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. Opcionalmente, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. Opcionalmente, el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. En algunos métodos, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb, y el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb.

En una modalidad, la suma total de los brazos de homología 5' y 3' del LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb.

En una modalidad, la secuencia de superposición se localiza en el extremo 3' del primer ácido nucleico y el extremo 5' de la segunda secuencia de ácido nucleico. En una modalidad, la secuencia de superposición se localiza en el extremo 5' de la primera secuencia de ácido nucleico y el extremo 3' de la segunda secuencia de ácido nucleico.

En una modalidad, el agente tipo nucleasa se expresa a partir de un constructo de expresión que comprende una secuencia de ácido nucleico que codifica una nucleasa, y en donde el ácido nucleico se enlaza operativamente a un promotor activo en la célula. En una modalidad, el agente tipo nucleasa se expresa a partir de un ARNm que codifica la nucleasa. En una modalidad, la nucleasa es una nucleasa de dedos de zinc (ZFN). En una modalidad, la nucleasa es una nucleasa efectora de tipo activador de la transcripción (TALEN). En una modalidad, la nucleasa es una meganucleasa.

En una modalidad, el agente tipo nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) y un ARN guía (ARNg). En una modalidad, la proteína Cas es Cas9.

En una modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son de una especie diferente. En otra modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son ácidos nucleicos humanos. En algunos métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son de una especie 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.

En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones,

5 un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. En algunos métodos, la integración de uno o más del primer, el segundo y el tercer insertos de ácido nucleico en el locus genómico objetivo resulta uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos.

10 En una modalidad, la integración del primer y/o segundo insertos en el locus genómico da como resultado una adición de una secuencia exógena en el locus genómico.

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

20 En otra modalidad, la integración del primer y/o segundo insertos de ácido nucleico en el locus genómico da como resultado una delección de una secuencia endógena en el locus genómico objetivo. En una modalidad, la delección de la secuencia endógena en el locus genómico objetivo 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, 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.

25 En algunos métodos, la célula es una célula humana. En otros métodos, la célula es una célula no humana. En algunos métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS). En algunos 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.

30 En una modalidad, la célula es una célula pluripotente. En otra modalidad, la célula pluripotente es una célula madre embrionaria (ES). En algunas modalidades, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En otra modalidad, la célula es una célula madre pluripotente inducida (iPS).

35 En una modalidad, el locus genómico objetivo está en el genoma de la célula. En otra modalidad, el locus genómico objetivo está en el ADN extracromosómico dentro de la célula.

40 En una modalidad, la célula es una célula de fibroblasto.

45 En algunos métodos, la célula es una célula no humana. En otros métodos, la célula es de un humano. En algunas modalidades la célula es una célula de mamífero. En otra modalidad, la célula de mamífero es de un roedor. En algunos casos, el roedor es un ratón, una rata, o un hámster.

En algunos de los métodos anteriores, la secuencia de nucleótidos superpuesta facilita el reclutamiento de la maquinaria de recombinación al locus genómico objetivo.

50 La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo mediante los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta, en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

55 En otro ejemplo, se proporciona un método para mejorar la recombinación homóloga en un locus genómico objetivo en una célula cargando una maquinaria de recombinación en un vector de transformación, que comprende introducir en la célula un primer ácido nucleico y un segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de nucleótidos superpuesta, y en donde la secuencia de nucleótidos superpuesta facilita el reclutamiento de la maquinaria de recombinación al locus genómico objetivo.

60 En un ejemplo, el método mejora la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el primer ácido nucleico se introduce sin el segundo ácido nucleico. En un ejemplo, el método mejora la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo. Algunos de tales métodos mejoran la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo en comparación con los métodos en los que el segundo ácido nucleico se introduce sin el primer ácido nucleico. Algunos de



tales métodos mejoran la recombinación homóloga del primer y el segundo ácido nucleico en el locus genómico objetivo. En un ejemplo, la mejora de la recombinación homóloga es al menos 0,5 veces, 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.

5

En una modalidad, la secuencia de superposición del primer ácido nucleico es homóloga a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, o 99,9 % idéntico a la secuencia de superposición del segundo ácido nucleico. En una modalidad, la secuencia de superposición del primer ácido nucleico es 100 % idéntica a esa secuencia de superposición del segundo ácido nucleico.

10

En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 5 kb. En una modalidad, la secuencia de superposición es de aproximadamente 5 kb a aproximadamente 10 kb. En una modalidad, la secuencia de superposición es de aproximadamente 10 kb a aproximadamente 15 kb. En una modalidad, la secuencia de superposición es de aproximadamente 15 kb a aproximadamente 20 kb. En una modalidad, la secuencia de superposición es de aproximadamente 20 kb a aproximadamente 25 kb. En una modalidad, la secuencia de superposición es de aproximadamente 25 kb a aproximadamente 30 kb. En una modalidad, la secuencia de superposición es de aproximadamente 30 kb a aproximadamente 35 kb. En una modalidad, la secuencia de superposición es de aproximadamente 35 kb a aproximadamente 40 kb. En una modalidad, la secuencia de superposición es de aproximadamente 40 kb a aproximadamente 45 kb. En una modalidad, la secuencia de superposición es de aproximadamente 45 kb a aproximadamente 50 kb. En una modalidad, la secuencia de superposición es de aproximadamente 50 kb a aproximadamente 55 kb. En una modalidad, la secuencia de superposición es de aproximadamente 55 kb a aproximadamente 60 kb. En una modalidad, la secuencia de superposición es de aproximadamente 60 kb a aproximadamente 65 kb. En una modalidad, la secuencia de superposición es de aproximadamente 65 kb a aproximadamente 70 kb.

15

20

25

En una modalidad, la secuencia de superposición es de al menos 5 kb. En una modalidad, la secuencia de superposición es de al menos 10 kb. En una modalidad, la secuencia de superposición es de al menos 15 kb. En una modalidad, la secuencia de superposición es de al menos 20 kb. En una modalidad, la secuencia de superposición es de al menos 25 kb. En una modalidad, la secuencia de superposición es de al menos 30 kb. En una modalidad, la secuencia de superposición es de al menos 35 kb. En una modalidad, la secuencia de superposición es de al menos 40 kb. En una modalidad, la secuencia de superposición es de al menos 45 kb. En una modalidad, la secuencia de superposición es de al menos 50 kb. En una modalidad, la secuencia de superposición es de al menos 55 kb. En una modalidad, la secuencia de superposición es de al menos 60 kb. En una modalidad, la secuencia de superposición es de al menos 65 kb. En una modalidad, la secuencia de superposición es de al menos 70 kb.

30

35

En una modalidad, el primer ácido nucleico es un vector de transformación 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 no comprende una secuencia de nucleótidos que sea homóloga al locus genómico excepto la secuencia de superposición.

40

En una modalidad, el segundo ácido nucleico es un segundo vector de transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3', y el primer ácido nucleico no comprende una secuencia de nucleótidos que sea homóloga al locus genómico excepto la secuencia de superposición.

45

En una modalidad, el primer ácido nucleico es un primer vector de transformación 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 transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'. En una modalidad, el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico son fragmentos superpuestos de un ácido nucleico contiguo.

50

En un ejemplo, el vector de transformación es de aproximadamente 1 kb a aproximadamente 2 kb. En un ejemplo, el vector de transformación es de aproximadamente 2 kb a aproximadamente 5 kb. En un ejemplo, el vector de transformación es de aproximadamente 5 kb a aproximadamente 10 kb.

55

El vector de transformación es un vector de transformación grande (LTVEC). En algunos métodos, el vector de transformación es un LTVEC que varía de aproximadamente 10 kb a aproximadamente 200 kb. En algunos métodos, el primer vector de transformación es un primer LTVEC y/o el segundo vector de transformación es un segundo LTVEC. En algunos métodos, el primer vector de transformación es un primer LTVEC que varía de aproximadamente 10 kb a aproximadamente 200 kb, y/o el segundo vector de transformación es un segundo LTVEC que varía de aproximadamente 10 kb a aproximadamente 200 kb. En una modalidad, el LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. Opcionalmente, el primer LTVEC es de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a

60

65

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, o de aproximadamente 150 kb a aproximadamente 200 kb. Opcionalmente, el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb. En algunos métodos, el primer LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb, y el segundo LTVEC es 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, o de aproximadamente 150 kb a aproximadamente 200 kb.

En una modalidad, la suma total de los brazos de homología 5' y 3' del LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb. En algunos métodos, la suma total de los brazos de homología 5' y 3' del primer LTVEC es de 10 kb a aproximadamente 200 kb, y la suma total de los brazos de homología 5' y 3' del segundo LTVEC es de 10 kb a aproximadamente 200 kb.

En una modalidad, la secuencia de superposición se localiza en el extremo 3' del primer ácido nucleico y el extremo 5' de la segunda secuencia de ácido nucleico. En una modalidad, la secuencia de superposición se localiza en el extremo 5' de la primera secuencia de ácido nucleico y el extremo 3' de la segunda secuencia de ácido nucleico.

En una modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son de una especie diferente. En otra modalidad, el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico son ácidos nucleicos humanos. En algunos métodos, el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son de una especie 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.

En algunos métodos, la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos. En algunos métodos, la integración de uno o más del primer, el segundo y el tercer insertos de ácido nucleico en el locus genómico objetivo resulta uno o más de: (a) una adición de una secuencia exógena en el locus genómico objetivo; (b) una delección de una secuencia endógena en el locus genómico objetivo; o (c) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de estos.

En una modalidad, la integración del primer y/o segundo insertos en el locus genómico da como resultado una adición de una secuencia exógena en el locus genómico.

En algunos ejemplos, la célula transformada comprende un ADN genómico que comprende el primer y el segundo inserto de ácido nucleico juntos que varían de aproximadamente 5 kb a aproximadamente 500 kb. En algunos métodos, la célula transformada 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 varía de aproximadamente 5 kb a aproximadamente 500 kb.

En otra modalidad, la integración del primer y/o segundo insertos de ácido nucleico en el locus genómico da como resultado una delección de una secuencia endógena en el locus genómico objetivo. En una modalidad, la delección de la secuencia endógena en el locus genómico objetivo 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, 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.

En algunos métodos, la célula es una célula humana. En otros métodos, la célula es una célula no humana. En algunos métodos, la célula es una célula pluripotente, una célula madre hematopoyética, una célula madre neuronal, o una célula de fibroblastos. Opcionalmente, la célula pluripotente es una célula madre embrionaria (ES) o una célula madre

pluripotente inducida (iPS). En algunos 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.

5 En una modalidad, la célula es una célula pluripotente. En otra modalidad, la célula pluripotente es una célula madre embrionaria (ES). En algunas modalidades, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En otra modalidad, la célula es una célula madre pluripotente inducida (iPS).

10 En una modalidad, el locus genómico objetivo está en el genoma de la célula. En otra modalidad, el locus genómico objetivo está en el ADN extracromosómico dentro de la célula.

10 En una modalidad, la célula es una célula de fibroblasto.

15 En una modalidad, la célula es una célula no humana. En otra modalidad, la célula es de un humano. En otra modalidad, la célula es una célula de mamífero. En otra modalidad, la célula de mamífero es de un roedor. En otra modalidad, el roedor es un ratón, una rata, o un hámster.

20 La descripción además proporciona métodos para producir un animal no humano de generación F0, que comprende: (a) introducir una célula ES no humana en un embrión huésped no humano, en donde la célula ES no humana se produjo mediante los métodos anteriores; y (b) gestar el embrión huésped no humano en una madre sustituta, en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación. Opcionalmente, el animal no humano es un ratón o una rata.

Breve descripción de las figuras

25 La Figura 1 proporciona un esquema para un evento de transformación doble genómico en donde se dirige una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de selección de higromicina. El casete de selección de higromicina se escinde por una nucleasa de dedos de zinc (ZFN) o por un complejo CRISPR/Cas y se dirige con dos vectores de transformación grandes que comprenden un casete de selección de neomicina y más de 280 kb de segmentos de genes variables de la cadena kappa de inmunoglobulina humana. Los vectores de transformación grandes comprenden cada uno una secuencia de superposición de aproximadamente 20 kb, que permite la recombinación homóloga entre los vectores de transformación grandes. El evento de transformación insertó con precisión los segmentos de genes variables de la cadena kappa de inmunoglobulina humana de ambos vectores de transformación en una sola etapa de transformación. Las ubicaciones de las diversas sondas usadas para confirmar el evento de transformación se muestran como rectángulos encerrados. La secuencia del ratón se representa por el sombreado diagonal, hacia arriba, la secuencia humana se representa por la ausencia de sombreado, y los sitios de recombinación y los casetes de selección se representan por el sombreado diagonal hacia abajo, discontinuo. El esquema no es a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

40 La Figura 2 proporciona un esquema para un solo evento de transformación en donde una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de selección de higromicina se dirige con un vector de transformación 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 usadas para confirmar el evento de transformación se muestran como rectángulos encerrados. La secuencia del ratón se representa por el sombreado diagonal, hacia arriba, la secuencia humana se representa por la ausencia de sombreado, y los sitios de recombinación y los casetes de selección se representan por el sombreado diagonal hacia abajo, discontinuo. El esquema no es a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

50 La Figura 3 proporciona un esquema para la transformación y destrucción de un casete de selección de higromicina con el uso de 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 higromicina. El esquema no es a escala.

55 La Figura 4 proporciona un esquema para un evento de transformación triple genómico en donde se dirige una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de selección de higromicina. El casete de selección de higromicina se escinde por una nucleasa de dedos de zinc (ZFN) o por un complejo CRISPR/Cas y se dirige con tres vectores de transformación grandes que comprenden un casete de selección de neomicina y aproximadamente 370 kb de segmentos de genes variables de la cadena kappa de inmunoglobulina humana. Los vectores de transformación grandes comprenden cada uno una secuencia de superposición de aproximadamente 20 kb a aproximadamente 60 kb, lo que permite la recombinación homóloga entre los vectores de transformación grandes. El evento de transformación insertó con precisión los segmentos de genes variables de la cadena kappa de inmunoglobulina humana de los tres vectores de transformación en una sola etapa de transformación. Las ubicaciones de las diversas sondas usadas para confirmar el evento de transformación se muestran como rectángulos encerrados. La secuencia del ratón se representa por el sombreado diagonal, hacia arriba, la secuencia humana se representa por la ausencia de sombreado, y los sitios de recombinación y los casetes de selección se representan por el sombreado diagonal hacia abajo, discontinuo. El esquema no es a escala y, por ejemplo, no refleja el número real de segmentos de genes variables.

Definiciones

5 Los términos "proteína", "polipéptido", y "péptido", usados indistintamente en la presente descripción, incluyen formas poliméricas de aminoácidos de cualquier longitud, que incluyen aminoácidos codificados y no codificados y aminoácidos modificados o derivados químicamente o bioquímicamente. Los términos además incluyen polímeros que se han modificados, tales como polipéptidos que tienen esqueletos de péptidos modificados.

10 Los términos "ácido nucleico" y "polinucleótido", usados indistintamente en la presente descripción, incluyen formas poliméricas de nucleótidos de cualquier longitud, que incluyen ribonucleótidos, desoxirribonucleótidos, o análogos o versiones modificadas de estos. 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 otros productos naturales, químicamente modificados, bioquímicamente modificados, no naturales, o bases de nucleótidos derivatizadas.

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

20 "Optimización en codones" generalmente incluye un proceso para modificar una secuencia de ácido nucleico para la expresión mejorada en células huésped particulares mediante el reemplazo de al menos un codón de la secuencia nativa con un codón que se usa con más frecuencia o máxima frecuencia en los genes de la célula huésped mientras que mantiene 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 procariota o eucariota dada, que incluye una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, un 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 "Base de datos de uso de codones." Estas tablas pueden adaptarse de varias maneras. Ver Nakamura y otros. (2000) Nucleic Acids Research 28:292. Los algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en un huésped particular además están disponibles (ver, por ejemplo, Gene Forge).

35 "Enlace operable" o "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 transcripcionales.

40 El término "célula pluripotente" o "célula madre pluripotente" incluye una célula indiferenciada 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 por mamífero (célula iPS).

45 El término "célula madre embrionaria" o "célula ES" incluye una célula totipotente o pluripotente derivada de embriones 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.

50 El término "célula madre pluripotente inducida" o "célula iPS" incluye una célula madre pluripotente que puede derivarse directamente de una célula adulta diferenciada. Las células iPS humanas pueden generarse mediante la introducción de conjuntos específicos de factores de reprogramación en una célula no pluripotente que puede incluir, por ejemplo, Oct3/4, factores de transcripción de la familia Sox (por ejemplo, Sox1, Sox2, Sox3, Sox15), factores de transcripción de la familia Myc (por ejemplo, c-Myc, 1-Myc, n-Myc), factores de transcripción de la familia tipo Krüppel (KLF) (por ejemplo, KLF1, KLF2, KLF4, KLF5), y/o factores de transcripción relacionados, tales como NANOG, LIN28, y/o Glis1. Las células iPS humanas además pueden generarse, por ejemplo, mediante el uso de los miARN, moléculas pequeñas que imitan las acciones de 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 además se caracterizan por su capacidad de propagarse indefinidamente bajo condiciones de cultivo *in vitro* adecuadas. Ver, por ejemplo, Takahashi y Yamanaka (Cell (2006) Vol. 126(4), páginas 663-676).

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

65 "Complementariedad" de los ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos nucleobase, forma enlaces de hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases complementarias en el ADN son típicamente A con T y C con G. En el ARN, son típicamente C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad

perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en donde cada base en el dúplex se une a una base complementaria mediante el emparejamiento de Watson-Crick. Complementaria "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 ocurre una unión suficiente entre las bases en las dos cadenas para formar un complejo híbrido estable en conjunto de las condiciones de hibridación (por ejemplo, concentración de sal y temperatura). Tales condiciones pueden predecirse mediante el uso de secuencias y cálculos matemáticos estándar para predecir la  $T_m$  (temperatura de fusión) de las cadenas hibridadas, o mediante la determinación empírica de  $T_m$  mediante el uso de métodos de rutina. La  $T_m$  incluye la temperatura a la cual una población de complejos de hibridación formados entre dos cadenas de ácido nucleico se desnaturaliza al 50 % (es decir, una población de moléculas de ácido nucleico bicatenario se disocia a la mitad en cadenas simples). A una temperatura inferior a la  $T_m$ , se favorece la formación de un complejo de hibridación, mientras que a una temperatura superior a la  $T_m$ , se favorece la fusión o separación de las cadenas en el complejo de hibridación. La  $T_m$  puede estimarse para un ácido nucleico que tiene un contenido conocido de G+C en una solución acuosa de NaCl 1 M mediante el uso de, por ejemplo,  $T_m=81,5+0,41 (\% \text{ G+C})$ , aunque otros cálculos conocidos de  $T_m$  tienen en cuenta las características estructurales del ácido nucleico.

"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 complementarias y enlaces de hidrógeno para producir un complejo de hibridación. Tales condiciones incluyen los componentes químicos y sus concentraciones (por ejemplo, sales, agentes quelantes, formamida) de una solución acuosa u orgánica que contiene los ácidos nucleicos, y la temperatura de la mezcla. Otros factores, tales como la duración del tiempo de incubación o las dimensiones de la cámara de reacción pueden contribuir al entorno. Ver, por ejemplo, Sambrook y otros, *Molecular Cloning, A Laboratory Manual*, 2da edición, páginas 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles desajustes entre bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y el 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 ( $T_m$ ) para los híbridos de ácidos nucleicos que tienen esas secuencias. Para las hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo, complementariedad sobre 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 (ver Sambrook y otros, supra, 11.7-11.8). Típicamente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable incluyen al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 22 nucleótidos, al menos aproximadamente 25 nucleótidos, y al menos aproximadamente 30 nucleótidos. Además, la temperatura y la concentración de sal de la solución de lavado pueden ajustarse según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

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

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

Los métodos y composiciones proporcionados en la presente descripción emplean una variedad de componentes diferentes. En toda la descripción se reconoce que algunos componentes pueden tener variantes y fragmentos activos. Tales componentes incluyen, por ejemplo, agentes tipo nucleasa, proteínas Cas, ARN de CRISPR, ARNtracr, y ARN guía. La actividad biológica para cada uno de estos componentes se describe en otra parte de la presente descripción.

"Identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias de polipéptidos hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean para una correspondencia máxima sobre una ventana de comparación especificada. Cuando se usa el porcentaje de identidad de secuencia en referencia a las proteínas, se reconoce que las posiciones de los residuos que no son idénticas frecuentemente difieren por sustituciones conservadoras de aminoácidos, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y por lo tanto no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservativas, 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 en tales sustituciones conservadoras tienen "similitud de secuencia" o "similitud." Los medios para hacer este ajuste son bien conocidos. Típicamente, esto implica calificar una sustitución conservadora como un desajuste parcial en lugar de uno completo, aumentando de ese modo la identidad de secuencia porcentual. 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, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

"Porcentaje de identidad de secuencia" incluye el valor determinado comparando dos secuencias alineadas óptimamente en una ventana de comparación, en donde la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, interrupciones) 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 mediante la determinación del número de posiciones en las que la misma base de ácido nucleico o residuo de aminoácido ocurre en ambas secuencias para producir 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.

A menos que se indique de cualquier otra forma, los valores de identidad/similitud de secuencia incluyen el valor obtenido con el uso de GAP Versión 10 con el uso de los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos con el uso de un peso de GAP de 50 y un peso de longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos con el uso de un peso de GAP de 8 y un peso de longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de estos. "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 nucleótidos o residuos de aminoácidos y una identidad de secuencia porcentual idéntica cuando se compara con la alineación correspondiente generada por GAP Versión 10.

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 70 %, al menos 75 %, al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 %, o 100 % idéntica a la secuencia de referencia conocida. 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.

El término "*in vitro*" incluye entornos artificiales y a 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 u organismo o cuerpo) y a procesos o reacciones que ocurren dentro de un entorno natural. El término "*ex vivo*" incluye células que se han eliminado del cuerpo de un individuo y a procesos o reacciones que ocurren dentro de tales células.

Las composiciones o métodos que "comprenden" o "que incluyen" uno o más elementos mencionados pueden incluir otros elementos no específicamente mencionados. Por ejemplo, una composición que "comprende" o "incluye" una proteína puede contener la proteína sola o en combinación con otros ingredientes.

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

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

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

Descripción detallada de la invención

#### I. Modificación de loci genómicos con el uso de múltiples vectores de transformación

Se proporcionan métodos y composiciones para modificar un locus genómico objetivo dentro de una célula. Tales métodos emplean múltiples vectores de transformación grandes (LTVEC) que son capaces de recombinarse entre sí para formar un solo segmento contiguo de ácido nucleico. Tales métodos pueden utilizar 1, 2, 3, 4, 5, 6 o más LTVEC en una sola etapa de transformación. Se describen métodos y composiciones para mejorar la recombinación homóloga en un locus genómico objetivo en una célula. Tales métodos emplean dos o más ácidos nucleicos que comprenden una o más secuencias de superposición. Cualquiera de los métodos descritos en la presente descripción puede ocurrir *in vitro*, *ex vivo*, o *in vivo*.

#### A. Transformación doble

Se proporcionan métodos y composiciones para modificar un locus genómico objetivo dentro de una célula mediante un método de transformación doble. Los métodos y composiciones emplean dos vectores de transformación grandes (LTVEC) (es decir, un primer LTVEC y un segundo LTVEC) que son capaces de recombinarse entre sí para formar un solo segmento contiguo de ácido nucleico. 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 se flanquean 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 superposición LTVEC (es decir, son complementarios entre sí) y el primer y el segundo inserto flanquean los brazos de homología superpuestos. Tales métodos implican 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 objetivo; y (3) recombinación entre el brazo de homología 3' del segundo LTVEC y el segmento correspondiente en el locus objetivo. Esta recombinación de tres vías reconstruye el ácido nucleico contiguo en el locus objetivo con la secuencia de superposición de los brazos de homología colocados entre el primer y el segundo inserto de ácido nucleico.

Cada uno de los LTVEC además comprende o un brazo de homología 5' o 3' que es homólogo a una región de ADN dentro o cerca de un locus genómico objetivo que permite la recombinación e integración del segmento de ácido nucleico contiguo solo. Así, por medio de un evento de recombinación de tres vías, puede realizarse una gran modificación de ácido nucleico (es decir, delección, inserción y/o reemplazo) en un locus objetivo en una sola etapa de transformación.

Los tres eventos de recombinación pueden ocurrir en cualquier orden. En una modalidad, el evento de recombinación entre las secuencias de superposición de los dos LTVEC ocurre antes de la recombinación homóloga con el locus objetivo. En otra modalidad, la recombinación con el locus objetivo ocurre antes de la recombinación entre los dos LTVEC. Aun en otra modalidad, los tres eventos de recombinación pueden ocurrir simultáneamente.

En una modalidad, se proporciona un método para modificar un locus genómico objetivo en una célula. Un método de ese tipo comprende introducir un primer vector de transformación 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 donde el primer brazo de homología 3' del primer LTVEC tiene una secuencia de superposición 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 objetivo; en donde el locus genómico objetivo se modifica mediante la integración del primer y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes. El método comprende además seleccionar una célula transformada que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo.

#### B. Transformación triple

Se proporcionan además métodos y composiciones para modificar un locus genómico objetivo dentro de una célula mediante métodos de transformación triple. Los métodos y composiciones emplean tres vectores de transformación grandes (LTVEC) (es decir, un primer LTVEC, un segundo LTVEC y un tercer LTVEC) que son capaces de recombinarse entre sí para formar un solo segmento contiguo de ácido nucleico. 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 se flanquean 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 superposición LTVEC (es decir, son complementarios entre sí) y el primer y el segundo inserto 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 el segundo y tercer inserto flanquean los brazos de homología superpuestos.

Tales 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 objetivo; y (4) recombinación entre el brazo de homología 3' del tercer LTVEC y el segmento correspondiente en el locus objetivo. Esta recombinación de cuatro vías reconstruye el ácido nucleico contiguo en el locus objetivo con la secuencia de superposición de los brazos de homología colocados entre el primer y el segundo inserto de ácido nucleico y entre el segundo y el tercer inserto de ácido nucleico.

El primer y el tercer LTVEC además comprenden un brazo de homología 5' o 3' que es homólogo a una región de ADN dentro o cerca de un locus genómico objetivo, que permite la recombinación e integración del segmento de ácido nucleico

contiguo solo. Así, por medio de un evento de recombinación de cuatro vías, puede realizarse una gran modificación de ácido nucleico (es decir, delección, inserción y/o reemplazo) en un lugar objetivo en una etapa de transformación sola.

5 Los cuatro eventos de recombinación pueden ocurrir en cualquier orden. En una modalidad, el evento de recombinación entre las secuencias de superposición de los tres LTVEC ocurre antes de la recombinación homóloga con el locus objetivo. En otra modalidad, la recombinación con el locus objetivo ocurre antes de la recombinación entre los tres LTVEC. Aun en otra modalidad, los cuatro eventos de recombinación pueden ocurrir simultáneamente.

10 En una modalidad, se proporciona un método para modificar un locus genómico objetivo en una célula. Un método de ese tipo comprende introducir un primer vector de transformación 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 donde el primer brazo de homología 3' del primer LTVEC tiene una secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una secuencia de superposición 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 objetivo; en donde el locus genómico objetivo se modifica mediante la integración del primer, segundo, y tercer insertos de ácido nucleico entre los segmentos genómicos correspondientes. El método comprende además seleccionar una célula transformada que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo.

#### 25 C. Transformación con múltiples LTVEC

Los métodos de transformación proporcionados en la presente descripción para crear una modificación genética en una sola etapa de transformación proporcionan nuevas posibilidades y eficiencias mejoradas para las modificaciones de genes dirigidos más allá de las logradas con un solo método de transformación LTVEC. La transformación 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 de superposición de los LTVEC puede ocurrir antes de la recombinación homóloga con el locus objetivo. Alternativamente, la recombinación con el locus objetivo puede ocurrir antes de la recombinación entre los LTVEC o los eventos de recombinación pueden ocurrir simultáneamente.

35 Los métodos de transformación descritos en la presente descripción brindan varias ventajas sobre los métodos de transformación de LTVEC individuales existentes, que incluyen una mayor eficiencia de transformación, un aumento en el tamaño alcanzable de la modificación genética, y una reducción en el número de etapas de transformación necesarios para obtener grandes modificaciones genómicas, lo que ahorra tiempo y mantiene la pluripotencia de las células madre embrionarias no humanas modificadas. Esto es de particular importancia para 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. Así, tales modificaciones pueden permitir delecciones, reemplazos e inserciones muy grandes (por ejemplo, >50 kb) dentro del locus genómico a transformar.

45 Por ejemplo, el tiempo requerido para usar tres LTVEC de manera secuencial para modificar un locus genómico objetivo y tamizar y confirmar la modificación dirigida es de aproximadamente nueve meses, mientras que la misma modificación puede hacerse y confirmar con tres LTVEC simultáneamente en solo cuatro meses.

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

60 En modalidades 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 con un ácido nucleico heterólogo, una desactivación, o una activación. En ejemplos específicos, la modificación genética se introduce mediante la introducción de al menos dos vectores de transformación grandes (LTVEC) en una célula. En otro ejemplo, la modificación genética se introduce al introducir al menos tres vectores de transformación grandes (LTVEC) en una célula. En tales ejemplos, los LTVEC pueden comprender ADN para insertarse en el locus genómico objetivo de la célula.

65



En algunas modalidades, los métodos para modificar un locus genómico objetivo comprenden introducir una modificación genética en células de mamífero. De manera similar, la invención proporciona células de mamífero que comprenden una modificación genética.

5 Pueden usarse 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 puede modificarse con el uso de agentes tipo nucleasa que generan una ruptura de simple o doble cadena en un locus genómico a transformar. La ruptura de simple o doble cadena se repara después por la vía de unión del extremo no homólogo (NHEJ). Métodos ilustrativos para generar tales modificaciones genéticas dirigidas se analizan en detalle en otra parte de la presente descripción, que incluyen, por ejemplo, el uso de vectores de transformación grandes. Ver además Wang y otros. (2013) Cell 153:910-918, Mandalos y otros. (2012) PLOS ONE 7:e45768:1-9, y Wang y otros. (2013) Nat Biotechnol. 31:530-532.

15 La modificación genética dirigida por recombinación homóloga entre un vector de transformación y un locus objetivo puede ser muy ineficiente, especialmente en tipos de células distintas de las células madre embrionarias de roedores. El uso de un vector de transformación en combinación con una ruptura de ADN bicatenario dirigida por nucleasas en el locus objetivo puede mejorar en gran medida la eficacia de la transformación para modificaciones, tales como deleciones o inserciones. Similarmente, el uso de un vector de transformación en combinación con una ruptura de ADN de cadena monocatenaria dirigida por nucleasa en el locus objetivo puede mejorar en gran medida la eficacia de la transformación para modificaciones.

25 En algunas modalidades, los LTVEC pueden emplearse en combinación con agentes tipo nucleasa que producen una ruptura de simple o doble cadena dentro de un locus genómico objetivo. Un método de ese tipo comprende además introducir un agente tipo nucleasa en una célula. En una modalidad, el agente tipo nucleasa es una nucleasa de dedos de zinc (ZFN). En otra modalidad, el agente tipo nucleasa es un sistema de repeticiones palindrómicas cortas agrupadas regularmente entrecruzadas (CRISPR)/asociadas a CRISPR (Cas).

30 En una modalidad, se proporciona un método para modificar un locus genómico objetivo en una célula utilizando múltiples LTVEC. Un método de ese tipo comprende (a) introducir en una célula un agente tipo nucleasa que produce una ruptura de una cadena monocatenaria o bicatenaria dentro de un locus genómico objetivo; (b) introducir un primer vector de transformación 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 donde 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 objetivo y el primer brazo de homología 3' del primer LTVEC y el segundo brazo 5' del segundo LTVEC son homólogos entre sí o respectivamente a otros brazos de homología 5' y 3' de uno o más LTVEC adicionales, cada uno de los cuales incluye un inserto de ácido nucleico adicional flanqueado por otro brazo de homología 5' y otro brazo de homología 3' adicional; en donde el locus genómico objetivo se modifica mediante la integración del primer inserto de ácido nucleico, uno o más insertos de ácido nucleico adicionales de uno o más LTVEC adicionales (si están presentes), y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula objetivo que comprende el primer inserto de ácido nucleico, uno o más insertos de ácido nucleico adicionales (si están presentes) y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo. En tales métodos, los LTVEC adicionales son uno o más LTVEC que, cuando están presentes, se insertan entre el primer y el segundo LTVEC.

45 En una modalidad, se proporciona un método de transformación doble para modificar un locus genómico objetivo en una célula, el método comprende: (a) introducir en una célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo; (b) introducir un primer vector de transformación 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 donde el primer brazo de homología 3' del primer LTVEC tiene una secuencia de superposición 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 es homólogo a los segmentos genómicos correspondientes dentro del locus genómico objetivo; en donde el locus genómico objetivo se modifica mediante la integración del primer y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula objetivo que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo. En tales 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 objetivo.

65 En una modalidad, se proporciona un método de transformación triple para modificar un locus genómico objetivo en una célula, el método comprende: (a) introducir en una célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro de un locus genómico objetivo; (b) introducir un primer vector de transformación 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 donde el primer brazo de homología 3' del primer LTVEC tiene una secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una secuencia de superposición 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 objetivo; en donde el locus genómico objetivo se modifica mediante la integración del primer, segundo y tercer insertos de ácido nucleico entre los segmentos genómicos correspondientes; y (c) seleccionar una célula objetivo que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo. En dichos métodos de transformación 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 objetivo.

En algunos casos, los dos, tres, o más LTVEC pueden introducir simultáneamente. Alternativamente, los dos, tres, o más LTVEC pueden introducirse de forma secuencial o en diferentes momentos.

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

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

Los métodos de transformación proporcionados en la presente descripción para crear una modificación genética en una sola etapa de transformación proporcionan nuevas posibilidades y eficiencias mejoradas para modificaciones de genes dirigidos más allá de las logradas con un solo ácido nucleico. La transformación con dos, tres, o más ácidos nucleicos que son capaces de recombinarse entre sí permite la modificación de un segmento más grande de ADN y puede proporcionar eficiencias de transformación mejoradas sobre ácidos nucleicos individuales solos, incluso en ausencia de agentes tipo nucleasa. Tales métodos sin agentes tipo nucleasas pueden ser ventajosos sobre aquellos que emplean agentes tipo nucleasas porque la detección requerida para los métodos que usan nucleasas es más complicada y requiere más tiempo, lo que implica las etapas de tamizaje adicionales para confirmar la escisión y verificar los efectos no dirigidos. Los ácidos nucleicos (por ejemplo, los LTVEC) con regiones superpuestas de longitud suficiente pueden mejorar la recombinación homóloga en un locus genómico objetivo incluso en ausencia de una nucleasa dirigida. Como ejemplo, el uso de dos ácidos nucleicos con una región solapada de longitud suficiente puede mejorar la recombinación homóloga en un locus genómico objetivo en comparación con el uso de un solo ácido nucleico. Aunque no se requiere una comprensión del mecanismo para la práctica, se cree que la recombinación homóloga se mejora en tales circunstancias mediante la carga de la maquinaria de recombinación (por ejemplo, Exo1, Rad51, BRCA2, y así sucesivamente) en los ácidos nucleicos (por ejemplo, LTVEC), facilitando de ese modo el reclutamiento de la maquinaria de recombinación al locus objetivo.

En esta descripción se proporcionan métodos para modificar un locus genómico objetivo o mejorar la recombinación homóloga en un locus genómico objetivo en una célula, que comprende introducir en la célula el primer y el segundo ácido nucleico, en donde el primer y el segundo ácido nucleico comprenden una secuencia de superposición. El primer y segundo ácidos nucleicos pueden ser, por ejemplo, ácidos nucleicos lineales. Tales métodos además pueden comprender la introducción en la célula de tres o más ácidos nucleicos que son capaces de recombinarse entre sí. Por ejemplo, el primer y segundo ácidos nucleicos pueden tener una primera secuencia de superposición, y el segundo y tercer ácido nucleico pueden tener una segunda secuencia de superposición. En algunos métodos, el locus genómico objetivo se modifica, o la recombinación homóloga en el locus genómico objetivo se potencia, sin la ayuda de una nucleasa. En otros métodos, se modifica el locus genómico objetivo, o se mejora la recombinación homóloga en el locus genómico objetivo, con la ayuda de una nucleasa que hace que una cadena monocatenaria o bicatenaria se rompa o se acerque al locus genómico objetivo, como una nucleasa de dedos de zinc, un TALEN, una meganucleasa, o Cas9 y un ARN guía.

El método de esta descripción puede potenciar la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo, puede potenciar la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo, o puede potenciar la recombinación homóloga tanto del primer como del segundo ácidos nucleicos en el locus genómico objetivo. Como un ejemplo, la recombinación homóloga del primer ácido nucleico en el locus genómico objetivo puede mejorarse en comparación con los métodos en el que el primer ácido nucleico se introduce sin el segundo ácido nucleico. De manera similar, la recombinación homóloga del segundo ácido nucleico en el locus genómico objetivo puede potenciarse 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 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 mejora sin una nucleasa puede ser comparable a la mejora con una nucleasa. Por ejemplo, el cambio en veces en la mejora con una nucleasa puede ser 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 nucleasa. En algunos casos, la mejora sin una nucleasa puede ser igual o mayor que la mejora con una nucleasa.

5 La secuencia de superposición del primer ácido nucleico puede ser homóloga a la secuencia de superposición del segundo ácido nucleico. Por ejemplo, la secuencia de superposición 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 de superposición del segundo ácido nucleico. Alternativamente, la secuencia de superposición del primer ácido nucleico puede ser 100 % idéntica a la secuencia de superposición del segundo ácido nucleico.

10 La secuencia de superposición puede ser, por ejemplo, de aproximadamente 1 kb a aproximadamente 70 kb o más. Por ejemplo, la secuencia de superposición 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 aproximadamente 280 kb a aproximadamente 300 kb. Como un ejemplo, la secuencia de superposición puede ser de aproximadamente 20 kb a aproximadamente 60 kb. Alternativamente, la secuencia de superposición puede ser 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 de superposición puede ubicarse en cualquier lugar dentro del primer y segundo ácido nucleico. Por ejemplo, la secuencia de superposición puede ubicarse en el extremo 3' del primer ácido nucleico y el extremo 5' del segundo ácido nucleico. Alternativamente, la secuencia de superposición puede ubicarse en el extremo 5' del primer ácido nucleico y en el extremo 3' del segundo ácido nucleico.

40 En algunos métodos, el primer ácido nucleico es un vector de transformación que comprende un primer inserto de ácido nucleico flanqueado por un primer brazo de homología 5' y un primer brazo de homología 3'. El segundo ácido nucleico puede ser cualquier ácido nucleico que comprende una secuencia de superposición, tal como un plásmido, un vector de transformación, o un vector de transformación grande. En algunos métodos, el segundo ácido nucleico no comprende una secuencia de nucleótidos que sea homóloga al locus genómico objetivo, excepto por la secuencia de superposición. Por ejemplo, el segundo ácido nucleico puede consistir esencialmente o consistir en la secuencia de superposición.

45 En algunos métodos, el primer ácido nucleico es un vector de transformación 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 transformación que comprende un segundo inserto de ácido nucleico flanqueado por un segundo brazo de homología 5' y un segundo brazo de homología 3'.

50 El primer vector de transformación de esta descripción puede ser de cualquier tamaño. De manera similar, el segundo vector de transformación de esta descripción puede ser de cualquier tamaño. Por ejemplo, el primer y/o segundo vector de transformación puede ser de aproximadamente 1 kb a aproximadamente 2 kb, de aproximadamente 2 kb a aproximadamente 5 kb, o de aproximadamente 5 kb a aproximadamente 10 kb. El primer vector de transformación además puede ser un vector de transformación grande (LTVEC). De manera similar, el segundo vector de transformación puede ser un LTVEC. Tamaños ilustrativos de los LTVEC se describen en otra parte de la presente descripción. Por ejemplo, el primer y/o segundo LTVE 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, 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.

65

En algunos métodos de esta descripción, el primer ácido nucleico es un LTVEC, y el segundo ácido nucleico es un ácido nucleico más pequeño que comprende una secuencia de superposición, tal como un plásmido o un vector de transformación. En algunos métodos, el segundo ácido nucleico no comprende una secuencia de nucleótidos que sea homóloga al locus genómico objetivo, excepto por la secuencia de superposición. Por ejemplo, el segundo ácido nucleico puede consistir esencialmente o consistir en la secuencia de superposición.

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, los insertos de ácido nucleico primero y/o segundo 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.

Los métodos pueden dar como resultado la integración del primer y/o segundo insertos de ácido nucleico en el locus genómico objetivo. La integración puede dar como resultado la adición de una secuencia en el locus genómico objetivo, la delección de una secuencia en el locus genómico objetivo, o el reemplazo de una secuencia en el locus genómico objetivo. Por ejemplo, la integración puede dar como resultado la adición de una secuencia exógena en el locus genómico objetivo, la delección de una secuencia endógena en el locus genómico objetivo, o el reemplazo de una secuencia endógena con una secuencia exógena en el locus genómico objetivo. El primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o la combinación del primer y el segundo inserto de ácido nucleico que se insertan en el locus genómico objetivo pueden ser, por ejemplo, de aproximadamente 5 kb a aproximadamente 500 kb. Otros insertos de ácido nucleico ilustrativos y tamaños de inserción se describen en otra parte de la presente descripción. La delección en el locus genómico objetivo puede ser, por ejemplo, de aproximadamente 5 kba aproximadamente 10 kb, de aproximadamente 10 kba aproximadamente 20 kb, de aproximadamente 20 kba aproximadamente 40 kb, de aproximadamente 40 kba aproximadamente 60 kb, de aproximadamente 60 kba aproximadamente 80 kb, de aproximadamente 80 kba aproximadamente 100 kb, de aproximadamente 100 kba aproximadamente 150 kb, o de aproximadamente 150 kba aproximadamente 200 kb, de aproximadamente 200 kba aproximadamente 300 kb, de aproximadamente 300 kba aproximadamente 400 kb, de aproximadamente 400 kba aproximadamente 500 kb, de aproximadamente 500 kba aproximadamente 600 kb, de aproximadamente 600 kba aproximadamente 700 kb, o de aproximadamente 700 kba aproximadamente 800 kb. Otros tamaños de delección ilustrativos se describen en otra parte de la presente descripción.

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

## II. Insertos de ácido nucleico y vectores de transformación

### A. Inserto de ácido nucleico

Pueden emplearse uno o más insertos de ácido nucleico en los métodos de esta descripción descritos en la presente descripción, y pueden introducirse en una célula a través de vectores de transformación separados o en el mismo vector de transformación. Los insertos de ácido nucleico incluyen segmentos de ADN que se integrarán en los loci objetivo genómicos. La integración de un inserto de ácido nucleico en un locus objetivo puede dar como resultado la adición de una secuencia de ácido nucleico de interés para el locus objetivo, la delección de una secuencia de ácido nucleico de interés en el locus objetivo, y/o el reemplazo de una secuencia de ácido nucleico de interés en el locus objetivo.

Los métodos proporcionan la modificación de un locus genómico con insertos de ácido nucleico que son más grandes en tamaño de lo que puede lograrse con el uso de técnicas convencionales de transformación solo (es decir, un LTVEC solo). En tales métodos, los insertos de ácido nucleico se incluyen en dos, tres, o más LTVEC. Los LTVEC se diseñan 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.

En tales métodos, los insertos de ácido nucleico se flanquean 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 después se reforma mediante recombinación entre los fragmentos superpuestos de los brazos de homología. En tales métodos, la recombinación entre los dos LTVEC da como resultado un inserto de ácido nucleico contiguo con la secuencia de superposición de los brazos de homología colocados entre el primer y el segundo inserto de ácido nucleico. Los métodos de transformación 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 después se reforma mediante recombinación entre los fragmentos superpuestos de los brazos de homología. En tales métodos de transformación triple, la recombinación entre los tres LTVEC da como resultado un inserto contiguo de ácido nucleico con la secuencia de superposición de los brazos de homología colocados entre el primer, el segundo, y el tercer inserto de ácido nucleico. En una modalidad, la secuencia de superposición de los brazos de homología comprende una porción del inserto de ácido nucleico.

Como tal, estos métodos permiten la modificación de un locus genómico con insertos de ácido nucleico de dos, tres o más LTVEC en una etapa de transformación sola, lo que aumenta efectivamente el tamaño total del inserto de ácido nucleico y al mismo tiempo reduce el número de etapas de transformación.

5 El inserto de ácido nucleico o el ácido nucleico correspondiente en el locus objetivo que se está reemplazando puede ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor, un potenciador, o cualquier combinación de estos. Además, en algunos ejemplos, el tamaño del inserto de ácido nucleico (es decir, los insertos de ácido nucleico combinados de los dos, tres o más LTVEC) o el ácido nucleico correspondiente en el locus objetivo que se reemplaza puede ser de cualquier longitud deseada, que incluyen, por ejemplo, entre 10-100 nucleótidos de longitud, 100-500 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, los insertos de ácido nucleico combinados de los dos, tres, o más LTVEC o el ácido nucleico correspondiente en el locus objetivo que se reemplaza puede 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 otros casos, la longitud puede ser de 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. Por ejemplo, los insertos de ácido nucleico combinados de dos, tres o más LTVEC) o el ácido nucleico correspondiente en el locus objetivo que se reemplaza puede ser 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 modalidad, el tamaño del inserto de ácido nucleico es de aproximadamente 5 kb a aproximadamente 700 kb. En una modalidad, el tamaño del inserto de ácido nucleico es de aproximadamente 5 kb a aproximadamente 500 kb. En otra modalidad, el tamaño del inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 700 kb. En otra modalidad, el tamaño del inserto de ácido nucleico es de aproximadamente 100 kb a aproximadamente 500 kb. En una modalidad específica, el inserto de ácido nucleico es de aproximadamente 140 kb. En otra modalidad específica, el inserto de ácido nucleico es de aproximadamente 370 kb. En otra modalidad específica, el inserto de ácido nucleico es de aproximadamente 300 kb. En otra modalidad específica, el inserto de ácido nucleico es de aproximadamente 400 kb.

En algunos vectores de transformación individuales (es decir, antes de la recombinación con otro vector de transformación), el inserto de ácido nucleico puede tener entre 10-100 nucleótidos de longitud, 100-500 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 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. 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, 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 casos, el reemplazo del ácido nucleico en el locus objetivo da como resultado la delección 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 delección es mayor que la longitud total del brazo de homología 5' y el brazo de homología 3'.

En algunos casos, la extensión de la delección 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 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 delección 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.

En otros casos, el inserto de ácido nucleico o el correspondiente ácido nucleico en el locus objetivo que se reemplaza puede ser 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 120 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 o más.

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 una procarionota, una eucariota, una levadura, un ave (por ejemplo, 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 para uso agrícola, o cualquier otro organismo de interés.

El inserto de ácido nucleico y/o el ácido nucleico en el locus objetivo pueden comprender una secuencia codificante o una secuencia no codificante, tal como un elemento regulador (por ejemplo, un promotor, un potenciador, o un elemento de unión al represor transcripcional). Por ejemplo, el inserto de ácido nucleico puede comprender un alelo de activación de al menos un exón de un gen endógeno, o un alelo de activación de todo el gen endógeno (es decir, "activación por intercambio de genes"). Por ejemplo, el inserto de ácido nucleico puede ser homólogo u ortólogo a una secuencia que se dirige a la delección en el locus genómico objetivo. El inserto de ácido nucleico homólogo u ortólogo puede reemplazar la secuencia que se dirige para la delección en el locus genómico de interés. Esto puede dar como resultado 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 homólogo u ortólogo (es decir, el inserto de ácido nucleico se inserta en lugar de la correspondiente secuencia de ADN humano en su locus genómico endógeno).

El inserto de ácido nucleico además 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 de sentido con respecto a la transcripción de un gen objetivo; (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 de trampa de gen invertible) en orientación inversa. Ver, por ejemplo, el documento US 2011/0104799. El alelo condicional puede comprender además unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de actuación y el DSC; y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido. Ver 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. Tal selección de marcadores incluye, pero no se limita a, neomicina, fosfotransferasa (*neo<sup>r</sup>*), higromicina B fosfotransferasa (*hyg<sup>r</sup>*), puromicina-N-acetiltransferasa (*puro<sup>r</sup>*), blastidina S desaminasa (*bsr<sup>r</sup>*), xantina/guanina fosforibosil transferasa (*gpt*), o la timidina quinasa del virus del herpes simple (HSV-tk), o una combinación de estos. El polinucleótido que codifica el marcador de selección puede estar operativamente unido a un promotor activo en una célula a la que se dirige. Ejemplos de promotores se describen en otra parte de la presente descripción.

En algunos vectores de transformación, el inserto de ácido nucleico comprende un gen reportero. Ejemplos de genes informadores son genes que codifican luciferasa, β-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), DsRoja, ZsVerde, MmGFP, mMorado, mCereza, tdTomate, mFresa, J-Roja, mNaranja, mKO, mCitrina, Venus, YPet, Esmeralda, CyPet, Cerulean, T-Zafiro, fosfatasa alcalina, y una combinación de estos. Tales genes informadores pueden estar unidos operativamente a un promotor activo en una célula a la que se dirige. Ejemplos de promotores se describen en otra parte de la presente descripción.

En algunos vectores de transformación, el inserto de ácido nucleico comprende uno o más casetes de expresión o casetes de delección. Un casete dado puede comprender una secuencia de nucleótidos de interés, un ácido nucleico que codifica un marcador de selección y/o un gen reportero, 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 de la presente descripción.

En algunos vectores de transformación, el inserto de ácido nucleico comprende un ácido nucleico flanqueado por secuencias objetivo de recombinación específicas del sitio. Aunque todo el inserto de ácido nucleico puede flanquearse por tales secuencias objetivo de recombinación específicas del sitio, cualquier región o polinucleótido individual de interés dentro del inserto de ácido nucleico además puede flanquearse por tales sitios. Las secuencias objetivo 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 estos. 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. A continuación de la integración del inserto de ácido nucleico en un locus dirigido, pueden eliminarse las secuencias entre los sitios de recombinación específicos del sitio.

#### B. Polinucleótidos de interés

Cualquier polinucleótido de interés puede estar contenido en varios insertos de ácido nucleico y de ese modo integrado en un locus genómico objetivo. Los métodos descritos en la presente descripción, proporcionan al menos 1, 2, 3, 4, 5, 6 o más polinucleótidos de interés para ser integrados en el locus genómico a transformar.

El polinucleótido de interés dentro del inserto de ácido nucleico cuando se integra en el locus genómico objetivo puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una delección de una secuencia de ácido nucleico endógeno y/o la adición de un polinucleótido exógeno o heterólogo u ortólogo en el locus genómico objetivo. En una modalidad, la modificación genética comprende un reemplazo de una secuencia de ácido nucleico endógeno con un polinucleótido exógeno de interés en el locus genómico objetivo. Así, los métodos proporcionados en la presente descripción permiten la generación de una modificación genética que comprende una desactivación, una delección, una inserción, un reemplazo ("activación"), una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de estos en un locus genómico objetivo. Tales modificaciones pueden ocurrir tras la integración del primer, segundo, tercer, cuarto, quinto, sexto, séptimo, o posteriores insertos de ácido nucleico en el locus genómico objetivo.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico objetivo 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 de una especie extraña, 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 de una especie extraña. "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 la especie). El polinucleótido de interés puede ser de cualquier organismo de interés, que incluyen, pero no se limita a, no humanos, 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. Así, el 1<sup>er</sup>2<sup>do</sup>3<sup>er</sup>4<sup>to</sup>5<sup>to</sup>6<sup>to</sup>7<sup>mo</sup>, y/o cualquiera de los siguientes insertos de ácido nucleico pueden comprender tales secuencias.

En una modalidad, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico objetivo es homólogo a un ácido nucleico humano. En otras modalidades adicionales, el polinucleótido de interés integrado en el locus objetivo es un fragmento de un ácido nucleico genómico. En una modalidad, 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 producto agrícola ácido nucleico de mamífero o un ácido nucleico de mamífero no agrícola o una de combinación de estos.

En una modalidad, el polinucleótido de interés puede variar 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 50 kb a 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 objetivo 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 incluye, por ejemplo, una secuencia reguladora, una secuencia promotora, una secuencia potenciadora, una secuencia de unión al represor transcripcional o una delección de una secuencia no codificante de proteínas, pero no comprende una delección de una secuencia codificante de proteínas. Además, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico objetivo 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 estos.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico objetivo puede comprender una modificación genética en una secuencia codificante. Tales modificaciones genéticas incluyen, pero no se limitan a, una mutación por delección 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 objetivo puede comprender un polinucleótido que codifica una proteína mutante. En una modalidad, la proteína mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada, y/o patrón de expresión alterado. En una modalidad, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus objetivo genómico comprende al menos un alelo de la 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, que incluyen, pero no se limita a, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un 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 objetivo además puede comprender una secuencia reguladora, que incluye, por ejemplo, una secuencia promotora, una secuencia potenciadora,



una secuencia de unión al represor transcripcional, o una secuencia terminadora transcripcional. En modalidades específicas, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico objetivo comprende un polinucleótido que tiene una delección de una secuencia que no codifica proteínas, pero no comprende una delección de una secuencia codificante de proteínas. En una modalidad, la delección de la secuencia que no codifica proteínas comprende una delección de una secuencia reguladora. En otra modalidad, la delección del elemento regulador comprende una delección de una secuencia promotora. En una modalidad, la delección del elemento regulador comprende una delección de una secuencia potenciadora. Un polinucleótido de ese tipo de interés puede provenir de cualquier organismo, que incluye, pero no se limita a, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un 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. Tales modificaciones dirigidas incluyen, pero no se limitan a, adiciones de uno o más nucleótidos, delecciones de uno o más nucleótidos, sustituciones de uno o más nucleótidos, una desactivación del polinucleótido de interés o una porción de este, una activación del polinucleótido de interés o una porción de este, un reemplazo de una secuencia de ácido nucleico endógeno con una secuencia de ácido nucleico heterólogo, o una combinación de estas. En modalidades 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 transformación

Los vectores de transformación pueden emplearse para introducir el inserto de ácido nucleico en un locus objetivo genómico y comprender el inserto de ácido nucleico y los brazos de homología que flanquean el inserto de ácido nucleico. Los vectores de transformación pueden estar en forma lineal o en forma circular, y pueden ser monocatenarios o bicatenarios. Los vectores de transformación pueden ser ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). Para facilitar la referencia, se refieren en la presente descripción a los brazos de homología como brazos de homología 5' y 3' (es decir, corriente arriba y corriente abajo). Esta terminología se refiere a la posición relativa de los brazos de homología con respecto al inserto de ácido nucleico dentro del vector de transformación. Los brazos de homología 5' y 3' corresponden a regiones dentro de un locus dirigido o a una región dentro de otro vector de transformación, que en la presente descripción se refieren como "secuencia 5' objetivo" y "secuencia 3' objetivo," respectivamente. En algunas modalidades, los brazos de homología pueden funcionar además como una secuencia objetivo 5' o una 3'.

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

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

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

Un brazo de homología y una secuencia objetivo o dos brazos de homología "corresponden" o son "correspondientes" entre sí cuando las dos regiones comparten entre sí un nivel suficiente de identidad de secuencia para actuar como sustratos de 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 objetivo determinada y el brazo de homología correspondiente que se encuentra en el vector de transformación (es decir, secuencia de superposición) o entre dos brazos de homología puede tener cualquier grado de identidad de secuencia que permita que se produzca la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología de un vector de transformación (o un fragmento de este) y una secuencia objetivo de otro vector de transformación o una secuencia objetivo de un locus genómico objetivo (o un fragmento de este) puede ser, al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de manera que las secuencias experimenten recombinación homóloga.

Además, una región de homología correspondiente entre un brazo de homología y la secuencia objetivo correspondiente puede ser de cualquier longitud que sea suficiente para promover la recombinación homóloga en el locus genómico objetivo. Por ejemplo, un brazo de homología determinado y/o una secuencia objetivo correspondiente pueden comprender regiones de homología correspondientes que son, al menos aproximadamente de 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 en otra parte de la presente descripción) de manera que un brazo de homología tenga homología suficiente para experimentar recombinación homóloga con las secuencias objetivo correspondientes dentro de un locus genómico objetivo de la célula o dentro de otro vector de transformación.

Las secuencias de superposición de un brazo de homología 3' de un primer vector de transformación y un brazo de homología 5' de un segundo vector de transformación o de un brazo de homología 3' de un segundo vector de transformación y un brazo de homología 5' de un tercer vector de transformación pueden ser de cualquier longitud que sea suficiente para promover la recombinación homóloga entre dichos vectores de transformación. Por ejemplo, una secuencia de superposición determinada de un brazo de homología puede comprender regiones superpuestas correspondientes que son al menos aproximadamente de 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 más, de manera que una secuencia de superposición del brazo de homología tenga homología suficiente para experimentar la recombinación homóloga con una secuencia de superposición correspondiente dentro de otro vector de transformación. En una modalidad, la secuencia de superposición es de 1-5 kb. En una modalidad, la secuencia de superposición es de aproximadamente 1 kb a aproximadamente 70 kb. En una modalidad, la secuencia de superposición es de aproximadamente 10 kb a aproximadamente 70 kb. En otra modalidad, la secuencia de superposición es de aproximadamente 10 kb a aproximadamente 50 kb. En una modalidad, la secuencia de superposición es de al menos 10 kb. En algunas modalidades, una secuencia de superposición es de al menos 20 kb. Por ejemplo, la secuencia de superposición puede ser de aproximadamente 1kb 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 aproximadamente 280 kb a aproximadamente 300 kb. Como un ejemplo, la secuencia de superposición puede ser de aproximadamente 20 kb a aproximadamente 60 kb. Alternativamente, la secuencia de superposición puede ser 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 que es nativo para una célula (por ejemplo, un locus dirigido), o alternativamente pueden corresponder a una región de un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula, que incluye, por ejemplo, transgenes, casetes de expresión, o regiones de ADN heterólogas o exógenas. Alternativamente, los brazos de homología pueden, en algunas modalidades, corresponder a una región en un vector de transformación en una célula. Los brazos de homología del vector de transformación pueden corresponder a una región de un cromosoma artificial de levadura (YAC), un cromosoma artificial de bacteria (BAC), un cromosoma artificial humano, o cualquier otra región diseñada por ingeniería genética contenida en una célula huésped adecuada. Aún más, los brazos de homología de un vector de transformación pueden corresponder a o derivarse de una región de una genoteca de BAC, una genoteca de cósmidos, o una genoteca de fagos PI. En modalidades determinadas, los brazos de homología de un vector de transformación corresponden a un locus que es nativo, heterólogo, o exógeno para un procarionte, una levadura, un ave (por ejemplo, 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 para uso agrícola, o cualquier otro organismo de interés. En algunos casos, los brazos de homología corresponden a un locus de la célula que no puede transformarse con el uso de un método convencional o que puede dirigirse solo de manera incorrecta o solo con una eficiencia significativamente baja en ausencia de una mella o ruptura de doble cadena inducida por un agente tipo nucleasa (por ejemplo, una proteína Cas). En algunas modalidades, los brazos de homología se derivan del ADN sintético.

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

En algunos vectores de transformación, 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 dirigido es un genoma de ratón de una primera cepa, y los brazos de transformación son de un genoma de ratón de una segunda cepa, en donde la primera cepa y la segunda cepa son diferentes. En modalidades determinadas, los brazos de homología son del genoma del mismo animal o son del genoma de la misma cepa, por ejemplo, el genoma dirigido es un genoma de ratón de una primera cepa, y los brazos de transformación son de un genoma de ratón del mismo ratón o de la misma cepa.

Un brazo de homología de un vector de transformación puede ser de cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con una secuencia objetivo 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 más. Como se describe detalladamente más abajo, los vectores de transformación grandes pueden emplear brazos de transformación de mayor longitud.

Los agentes tipo nucleasa (por ejemplo, sistemas CRISPR/Cas) pueden emplearse en combinación con los vectores de transformación para facilitar la modificación de un locus objetivo. Tales agentes tipo nucleasa pueden promover la recombinación homóloga entre un vector de transformación y un locus objetivo. Cuando los agentes tipo nucleasa se emplean en combinación con un vector de transformación, el vector de transformación puede comprender brazos de homología 5' y 3' correspondientes a las secuencias objetivo 5' y 3' ubicadas en proximidad suficiente a un sitio de escisión de la nucleasa para promover la ocurrencia de un evento de recombinación homóloga entre las secuencias objetivo y los brazos de homología tras una mella o ruptura de doble cadena en el sitio de escisión de la nucleasa. El término "sitio de escisión de la nucleasa" incluye una secuencia de ADN en la que se crea una mella o ruptura de doble cadena por un agente tipo nucleasa (por ejemplo, un sitio de escisión de Cas9). Las secuencias objetivo dentro de un locus objetivo que corresponden a los brazos de homología 5' y 3' de un vector de transformación se encuentran "ubicadas en proximidad suficiente" a un sitio de escisión de la nucleasa si la distancia es tal que promueve la ocurrencia de un evento de recombinación homóloga entre las secuencias objetivo 5' y 3' y los brazos de homología tras una mella o ruptura de doble cadena en el sitio de reconocimiento. Por lo tanto, en modalidades específicas, las secuencias objetivo correspondientes a los brazos de homología 5' y/o 3' de un vector de transformación están dentro de 1 nucleótido de un sitio de reconocimiento determinado o están dentro de al menos 10 nucleótidos hasta aproximadamente 14 kb de un sitio de reconocimiento determinado. En algunas modalidades, un sitio de escisión de la nucleasa está inmediatamente adyacente al menos a una o ambas de las secuencias objetivo.

La relación espacial de las secuencias objetivo que corresponden a los brazos de homología de un vector de transformación y el sitio de escisión de la nucleasa puede variar. Por ejemplo, las secuencias objetivo pueden ubicarse 5' a un sitio de escisión de la nucleasa, las secuencias objetivo pueden localizarse 3' a un sitio de reconocimiento, o las secuencias objetivo pueden flanquear el sitio de escisión de la nucleasa.

El uso combinado de un vector de transformación (que incluye, por ejemplo, un vector de transformación grande) con un agente tipo nucleasa puede dar como resultado una mayor eficiencia de transformación en comparación con el uso de un vector de transformación solo. Por ejemplo, cuando un vector de transformación se usa junto con un agente tipo nucleasa, la eficiencia de transformación de un vector de transformación puede aumentar al menos dos veces, al menos tres veces, al menos 4 veces, al menos 10 veces o dentro de un intervalo formado a partir de estos números enteros, tales como 2-10 veces en comparación con el uso del vector de transformación solo.

#### D. Vectores de transformación grandes

Algunos vectores de transformación son "vectores de transformación grandes" o "LTVEC", lo que incluye vectores de transformación que comprenden brazos de homología que corresponden a y se derivan de secuencias de ácidos nucleicos más grandes que las usadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en las células. Un LTVEC puede ser, por ejemplo, de al menos 10 kb de longitud, o la suma total del brazo de homología 5' y un brazo de homología 3' puede ser, por ejemplo, al menos 10 kb. Los LTVEC incluyen, además, vectores de transformación que comprenden insertos de ácido nucleico que tienen secuencias de ácidos nucleicos mayores que las usados típicamente por otros enfoques destinados a realizar la recombinación homóloga en las células. Por ejemplo, los LTVEC hacen posible la modificación de loci grandes que no pueden acomodarse por vectores de transformación tradicionales basados en plásmidos debido a sus limitaciones de tamaño. Por ejemplo, el locus objetivo puede ser (es decir, los brazos de homología 5' y 3' pueden corresponder a) un locus de la célula que no puede transformarse con el uso de un método convencional o que puede dirigirse solo de manera incorrecta o solo con una eficiencia significativamente baja en ausencia de una mella o ruptura de doble cadena inducida por un agente tipo nucleasa (por ejemplo, una proteína Cas).

Los métodos descritos en la presente descripción emplean dos o tres LTVEC que son capaces de recombinarse entre sí y con un locus genómico objetivo en un evento de recombinación de tres vías o cuatro vías como se describe en la presente en otro sitio. Tales métodos hacen posible la modificación de loci grandes que no puede lograrse con el uso de un solo LTVEC.

Los ejemplos de LTVEC incluyen vectores derivados de un cromosoma artificial de bacteria (BAC), un cromosoma artificial humano, o un cromosoma artificial de levadura (YAC). Se describen ejemplos de LTVEC y métodos para hacerlos, por ejemplo, en la patente de Estados Unidos núm. 6,586,251; patente de Estados Unidos núm. 6,596,541; patente de Estados Unidos núm. 7,105,348; y los documentos WO 2002/036789 (PCT/US01/45375). Los LTVEC pueden estar en forma lineal o en forma circular.

Los LTVEC de los ejemplos pueden ser de cualquier longitud, que incluyen, 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 ser al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más. El tamaño de un LTVEC puede ser demasiado grande para permitir el tamizaje de eventos de transformación mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de largo alcance (por ejemplo, 1 kb a 5 kb).

En algunos casos, un LTEVEC comprende un inserto de ácido nucleico en el intervalo 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 estar en el intervalo 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 LTVEC, la suma total del brazo de homología 5' y el brazo de homología 3' en un LTVEC es al menos 10 kb. En otros LTVEC, el brazo de homología 5' puede estar en el intervalo de aproximadamente 1 kb a aproximadamente 100 kb y/o el brazo de homología 3' puede estar en el intervalo 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 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 ser 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. Igualmente, la suma total de los brazos de homología 5' y 3' puede ser 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 LTEVEC y el inserto de ácido nucleico se diseñan para permitir una delección de la secuencia endógena en el locus genómico objetivo 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, 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 delección 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 delección puede ser 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 de 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 se diseñan para permitir una inserción en el locus objetivo de una secuencia de ácido nucleico exógena en el intervalo 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 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 de 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.

Aun en otros casos, el inserto de ácido nucleico y/o la región del locus endógeno que se elimina es 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 objetivo

Un genoma o locus genómico objetivo modificado por los métodos descritos en la presente descripción puede incluir cualquier segmento o región de ADN dentro de una célula. El genoma o locus genómico objetivo 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 estos. Dichos segmentos de ADN heterólogos o exógenos pueden incluir transgenes, casetes de expresión, marcadores de selección que codifican polinucleótidos o regiones heterólogas o exógenas del ADN genómico.

El genoma o locus genómico objetivo puede incluir además un ADN extracromosomal dentro de la célula, tal como un cromosoma artificial de levadura (YAC), un cromosoma artificial de bacteria (BAC), un cromosoma artificial humano, o cualquier otra región genómica diseñada por ingeniería genética contenida en una célula huésped adecuada.

#### III. Agentes tipo nucleasa

Los métodos y composiciones para modificar un locus genómico objetivo proporcionados en la presente descripción pueden emplear un agente tipo nucleasa que induce una mella o ruptura de doble cadena en un sitio de reconocimiento deseado.

El término "sitio de reconocimiento para un agente tipo nucleasa" incluye una secuencia de ADN en la que se induce una mella o ruptura por un agente tipo nucleasa. El sitio de reconocimiento para un agente tipo nucleasa puede ser endógeno (o nativo) a la célula o el sitio de reconocimiento puede ser exógeno a la célula. En modalidades específicas, el sitio de reconocimiento es exógeno a la célula y, por lo tanto, no ocurre naturalmente en el genoma de la célula. En otras modalidades adicionales, el sitio de reconocimiento es exógeno a la célula y a los polinucleótidos de interés que uno deseaba posicionar en el locus objetivo. En modalidades adicionales, el sitio de reconocimiento exógeno o endógeno está presente solo una vez en el genoma de la célula huésped. En modalidades específicas, se identifica un sitio endógeno o nativo que ocurre solo una vez dentro del genoma. Tal sitio puede utilizarse para diseñar agentes tipo nucleasa que producirán una mella o ruptura de doble cadena en el sitio de reconocimiento endógeno.

La longitud del sitio de reconocimiento puede variar e incluir, por ejemplo, sitios de reconocimiento que son de aproximadamente de 30-36 pb para un par de nucleasas de dedos de zinc (ZFN) (es decir, aproximadamente 15-18 pb para cada ZFN), aproximadamente 36 pb para una nucleasa de actividad similar al activador de la transcripción (TALEN), o aproximadamente 20 pb para un ARN guía CRISPR/Cas9.

5

Cualquier agente tipo nucleasa que induzca una mella o ruptura de doble cadena en un sitio de reconocimiento deseado puede utilizarse en los métodos y composiciones descritos en la presente. Se puede emplear un agente tipo nucleasa de origen natural o nativo siempre que el agente tipo nucleasa induzca una mella o ruptura de doble cadena en un sitio de reconocimiento deseado. Alternativamente, puede emplearse un agente tipo nucleasa modificado o diseñado por ingeniería genética. Un "agente tipo nucleasa diseñado por ingeniería genética" incluye una nucleasa que se diseña (modifica o deriva) de su forma nativa para reconocer e inducir específicamente una mella o ruptura de doble cadena en el sitio de reconocimiento deseado. Por lo tanto, un agente tipo nucleasa diseñado por ingeniería genética puede derivarse de un agente tipo nucleasa nativo, de origen natural o puede crearse o sintetizarse artificialmente. La modificación del agente tipo 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 modalidades, la nucleasa diseñada por ingeniería genética induce una mella o ruptura de doble cadena en un sitio de reconocimiento, en donde el sitio de reconocimiento no era una secuencia que hubiera sido reconocida por un agente tipo nucleasa nativo (sin diseñar por ingeniería genética o sin modificar). La producción de una mella o ruptura de doble cadena en un sitio de reconocimiento u otro ADN puede denominarse en la presente descripción "cortar" o "escindir" el sitio de reconocimiento u otro ADN.

20

Se proporcionan además variantes activas y fragmentos de los sitios de reconocimiento ejemplificados. Dichas variantes activas pueden comprender al menos 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más identidad de secuencia para el sitio de reconocimiento dado, en donde las variantes activas retienen actividad biológica y, por lo tanto, son capaces de ser reconocidas y escindidas por un agente tipo nucleasa de una manera específica de secuencia. Los ensayos para medir la ruptura de doble cadena de un sitio de reconocimiento por un agente tipo nucleasa son conocidos en la técnica (por ejemplo, el ensayo TAQMAN® qPCR, Frenthewey D. y otros, *Methods in Enzymology*, 2010, 476: 295-307).

25

El sitio de reconocimiento del agente tipo nucleasa puede colocarse en cualquier lugar dentro o cerca del locus objetivo. El sitio de reconocimiento puede localizarse 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 tipo nucleasa puede ubicarse en un intrón, un exón, un promotor, un potenciador, una región reguladora o cualquier región codificante no proteica. En modalidades 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 tipo nucleasa puede localizarse en un intrón del marcador de selección, un promotor, un potenciador, una región reguladora o cualquier región no codificante de proteínas del polinucleótido que codifica el marcador de selección. En modalidades específicas, una mella o ruptura de doble cadena en el sitio de reconocimiento interrumpe la actividad del marcador de selección. Se conocen métodos para analizar la presencia o ausencia de un marcador de selección funcional.

40

En una modalidad, el agente tipo nucleasa es una nucleasa de actividad similar al activador de la transcripción (TALEN). Las nucleasas efectoras TAL son una clase de nucleasas específicas de secuencia que pueden utilizarse para realizar rupturas de doble cadena en secuencias objetivo específicas en el genoma de un organismo procarionta o eucariota. Las nucleasas efectoras TAL se crean mediante la fusión de un efector similar al activador de la transcripción (TAL) nativo o diseñado por ingeniería genética, o parte funcional de este, al dominio catalítico de una endonucleasa, tal como, por ejemplo, FokI. El dominio modular único de unión al ADN efector TAL permite el diseño de proteínas con cualquier especificidad potencial de reconocimiento de ADN dado. Por lo tanto, los dominios de unión al ADN de las nucleasas efectoras de TAL pueden diseñarse por ingeniería genética para reconocer sitios específicos de ADN específicos y, por lo tanto, pueden utilizarse para realizar rupturas de doble cadena en las secuencias deseadas. Ver el documento WO 2010/079430; Morbitzer y otros, (2010) PNAS 10.1073/pnas.1013133107; Scholze y Boch (2010) *Virulence* 1: 428-432; Christian y otros, *Genética* (2010) 186:757-761; Li y otros, (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704; y Miller y otros, (2011) *Nature Biotechnology* 29:143-148.

45

50

Se describen ejemplos de nucleasas TAL adecuadas y métodos para preparar nucleasas TAL adecuadas, por ejemplo, en los documentos US 2011/0239315 A1, US 2011/0269234 A1, US 2011/0145940 A1, US 2003/0232410 A1, US 2005/0208489 A1, US 2005/0026157 A1, US 2005/0064474 A1, US 2006/0188987 A1 y US 2006/0063231 A1. En diversas modalidades, las nucleasas efectoras TAL se diseñan por ingeniería genética para cortar en o acercarse a una secuencia de ácido nucleico objetivo en, por ejemplo, un locus de interés o un locus genómico de interés, en donde la secuencia de ácido nucleico objetivo está en o cerca de una secuencia que se modifica por un vector de transformación. Las nucleasas TAL adecuadas para su uso con los diversos métodos y composiciones proporcionados en la presente descripción incluyen aquellas que están diseñadas específicamente para unirse a o acercarse a secuencias de ácido nucleico objetivo que se modifican por vectores de transformación como se describe en la presente descripción.

60

En una modalidad, cada monómero de TALEN comprende 33-35 repeticiones TAL que reconocen un solo par de bases a través de dos residuos hipervariables. En una modalidad, el agente tipo nucleasa es una proteína quimérica que comprende un dominio de unión al ADN basado en la repetición TAL operativamente enlazado a una nucleasa

65

independiente. En una modalidad, la nucleasa independiente es una endonucleasa FokI. En una modalidad, el agente tipo nucleasa comprende un primer dominio de unión al ADN basado en la repetición de TAL y un segundo dominio de unión al ADN basado en la repetición TAL, en donde cada uno de los primeros y segundos dominios de unión al ADN basado en la repetición TAL está operativamente enlazado a una nucleasa FokI, en donde el primer y el segundo dominios de unión a ADN basado en la repetición TAL reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separadas por una secuencia espaciadora de longitud variable (12-20 pb), y en donde las subunidades de la nucleasa FokI se dimerizan para crear una nucleasa activa que realiza una ruptura de doble cadena en una secuencia objetivo.

El agente tipo nucleasa empleado en los diversos métodos y composiciones descritos en la presente descripción puede comprender además una nucleasa de dedos de zinc (ZFN). En una modalidad, cada monómero de ZFN comprende 3 o más dominios de unión al ADN basados en dedos de zinc, en donde cada dominio de unión al ADN basado en dedos de zinc se une a un subsitio de 3 pb. En otras modalidades, el agente tipo nucleasa es una proteína quimérica que comprende un dominio de unión al ADN basado en la repetición TAL operativamente enlazado a una nucleasa independiente. En una modalidad, la endonucleasa independiente es una endonucleasa FokI. En una modalidad, el agente tipo nucleasa comprende una primera ZFN y una segunda ZFN, en donde cada una de la primera ZFN y segunda ZFN están operativamente enlazado a una subunidad de la nucleasa FokI, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separada por un espaciador de aproximadamente 5-7 pb, y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que realiza la ruptura de una doble cadena. Ver, por ejemplo, los documentos US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; y Gaj y otros, (2013) Trends in Biotechnology, 31(7): 397-405.

Todavía en otra modalidad, el agente tipo nucleasa es una meganucleasa. Las meganucleasas se han clasificado en cuatro familias basado en los motivos de la secuencia conservada, las familias son las familias de cajas LAGLIDADG, GIY-YIG, H-N-H, e His-Cys. Estos motivos participan en la coordinación de iones metálicos y la hidrólisis de enlaces fosfodiéster. Las meganucleasas son notables por sus sitios de reconocimiento largos y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Se conocen los dominios de meganucleasa, estructura y función, ver por ejemplo, Guhan y Muniyappa (2003) Crit Rev Biochem Mol Biol 38:199-248; Lucas y otros, (2001) Nucleic Acids Res 29:960-9; Jurica y Stoddard, (1999) Cell Mol Life Sci 55:1304-26; Stoddard, (2006) Q Rev Biophys 38:49-95; y Moure y otros, (2002) Nat Struct Biol 9:764. En algunos ejemplos, se usa una variante de origen natural y/o meganucleasa derivada diseñada por ingeniería genética. Se conocen los métodos para modificar la cinética, las interacciones de cofactores, la expresión, las condiciones óptimas y/o la especificidad del sitio de reconocimiento, y el tamizaje de la actividad, ver, por ejemplo, Epinat y otros, (2003) Nucleic Acids Res 31:2952-62; Chevalier y otros, (2002) Mol Cell 10:895-905; Gimble y otros, (2003) Mol Biol 334:993-1008; Seligman y otros, (2002) Nucleic Acids Res 30:3870-9; Sussman y otros, (2004) J Mol Biol 342:31-41; Rosen y otros, (2006) Nucleic Acids Res 34:4791-800; Chames y otros, (2005) Nucleic Acids Res 33:e178; Smith y otros, (2006) Nucleic Acids Res 34:e149; Gruen y otros, (2002) Nucleic Acids Res 30:e29; Chen y Zhao, (2005) Nucleic Acids Res 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.

Cualquier meganucleasa puede utilizarse en la presente descripción, incluyendo, pero sin limitarse a, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TII, I-PpoI, PI-Pspl, F-SceI, F-SceII, F-SuVI, F-TeV, F-TeVII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I-CvuAIIP, I-Ddil, I-DdIII, I-Dirl, 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-PbolIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PoblIP, I-PorI, I-PorIIP, I-PbplIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SnelIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelIP, I-TeV, I-TeVII, I-TeVIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbilIP, PI-MtuI, PI-MtuHIIP, PI-MtuHIIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-Tful, PI-Tfull, PI-Thyl, PI-TIil, PI-TIiil, o cualquiera de las variantes o fragmentos activos de esta.

En una modalidad, la meganucleasa reconoce secuencias de ADN de doble cadena de 12 a 40 pares de bases. En una modalidad, la meganucleasa reconoce una secuencia objetivo perfectamente coincidente en el genoma. En una modalidad, la meganucleasa es una nucleasa doméstica. En una modalidad, la nucleasa doméstica es una familia LAGLIDADG de la nucleasa doméstica. En una modalidad, la familia LAGLIDADG de la nucleasa doméstica se selecciona de I-SceI, I-CreI e I-Dmol.

Los agentes tipo nucleasa pueden comprender además endonucleasas de restricción, que incluyen endonucleasas de Tipo I, Tipo II, Tipo III y Tipo IV. Las endonucleasas de restricción de Tipo I y Tipo III reconocen sitios de reconocimiento específicos, pero típicamente se escinden en una posición variable del sitio de unión de la nucleasa, que pueden estar a cientos de pares de bases lejos 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 ocurre típicamente en los sitios específicos dentro o cerca del sitio de unión. La mayoría de las enzimas Tipo II cortan las secuencias palindrómicas, sin embargo, las enzimas Tipo IIa reconocen los sitios de reconocimiento no palindrómicos y escinden fuera del sitio de reconocimiento, las enzimas Tipo IIb cortan las secuencias dos veces con ambos sitios fuera del sitio de reconocimiento, y las enzimas Tipo IIc reconocen un sitio de reconocimiento asimétrico y escinden en 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 más detalladamente, por ejemplo, en la base de datos REBASE (página web en [rebase.neb.com](http://rebase.neb.com); Roberts y otros, (2003) *Nucleic Acids Res* 31:418-20), Roberts y otros, (2003) *Nucleic Acids Res* 31:1805-12y Belfort y otros, (2002) en *Mobile DNA II*, páginas 761-783, Eds. Craigie y otros, (ASM Press, Washington, DC).

5

El agente tipo nucleasa empleado en los diversos métodos y composiciones puede comprender además un sistema (CRISPR)/asociadas a CRISPR (Cas) de repeticiones palindrómicas cortas agrupadas regularmente intercaladas o componentes de dicho sistema. Los sistemas CRISPR/Cas incluyen transcritos y otros elementos implicados en la expresión de, o que direccionan la actividad de, los genes Cas. Un sistema CRISPR/Cas puede ser, por ejemplo, un sistema tipo I, un tipo II, o un tipo III. Los métodos y composiciones descritos en la presente descripción emplean sistemas CRISPR/Cas utilizando complejos CRISPR (que comprenden un ARN guía (ARNg) complejo con una proteína Cas) para la escisión dirigida al sitio de los ácidos nucleicos.

10

Algunos sistemas CRISPR/Cas usados en los métodos descritos en la presente descripción son de origen no natural. Un sistema "de origen no natural" incluye cualquier cosa que indique la participación de la mano del hombre, tal como uno o más componentes del sistema que están alterados o mutados de su estado natural, que están al menos sustancialmente libres 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 de origen no natural que comprenden un ARNg y una proteína Cas que no ocurren naturalmente juntos.

15

20

Las proteínas Cas generalmente comprenden al menos un dominio de reconocimiento o unión de ARN. Dichos dominios pueden interactuar con los ARN guía (ARNg, descritos con más detalle más abajo). Las proteínas Cas pueden comprender además dominios de nucleasa (por ejemplo, dominios de ADNasa o RNasa), dominios de unión al ADN, dominios helicasa, dominios de interacción proteína-proteína, dominios de dimerización y otros dominios. Un dominio de nucleasa posee actividad catalítica para la escisión del ácido nucleico. 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 en cadena simple o doble cadena. Una proteína Cas puede tener una actividad de escisión completa y crear una ruptura de doble cadena en un locus genómico objetivo (por ejemplo, una ruptura de doble cadena con extremos romos), o puede ser una nickasa que crea una ruptura de cadena simple en un locus genómico objetivo.

25

30

Los ejemplos de las 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 sus homólogos o versiones modificadas.

35

En algunos casos, una proteína Cas es de un sistema CRISPR/Cas 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.*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulfuridis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, o *Acaryochloris marina*. Las proteínas Cas9 pueden ser también 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 él. La secuencia de aminoácidos de una proteína Cas9 de *S. pyogenes* puede encontrarse, por ejemplo, en la base de datos SwissProt bajo el número de acceso Q99ZW2.

40

45

50

55

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, las que son de origen natural), proteínas Cas modificadas (es decir, variantes de la proteína Cas) o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las proteínas Cas pueden ser además variantes o fragmentos activos de proteínas Cas modificadas o de tipo silvestre. 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 para la proteína Cas de tipo silvestre o modificada o una porción de esta, en donde las variantes activas retienen la capacidad de cortar en un sitio de escisión deseado y, por lo tanto, retienen la actividad inductora de mella o ruptura de la doble cadena. Los ensayos para la actividad inductora de mella o ruptura de doble cadena se conocen y generalmente miden la actividad general y la especificidad de la proteína Cas en los sustratos de ADN que contienen el sitio de escisión.

60

65



Las proteínas Cas pueden modificarse para aumentar o disminuir la afinidad de unión al ácido nucleico, la especificidad de unión al ácido nucleico y/o la actividad enzimática. Las proteínas Cas pueden modificarse además para cambiar cualquier otra actividad o propiedad de la proteína, tal como la estabilidad. Por ejemplo, uno o más dominios de la nucleasa de la proteína Cas pueden modificarse, eliminarse o inactivarse, o una proteína Cas puede truncarse para eliminar los 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 similar a RuvC y un dominio de nucleasa similar a HNH. Los dominios RuvC y HNH pueden cortar una cadena diferente de ADN de doble cadena para realizar una ruptura de doble cadena en el ADN. Ver, por ejemplo, Jinek y otros. (2012) Science 337:816-821.

Uno o ambos dominios de nucleasa pueden eliminarse o mutar para que ya no sean funcionales o tengan actividad de nucleasa reducida. Si uno de los dominios de nucleasa se elimina o muta, la proteína Cas resultante (por ejemplo, Cas9) puede denominarse como una nickasa y puede generar una ruptura de simple cadena en una secuencia objetivo 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 (por ejemplo, Cas9) tendrá una capacidad reducida para escindir ambas cadenas de un ADN bicatenario (por ejemplo, una proteína Cas nula de nucleasa). Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato a alanina en la posición 10 de Cas9) en el dominio RuvC de Cas9 de *S. pyogenes*. De manera similar, H939A (histidina a alanina en la posición de aminoácidos 839) o H840A (histidina a alanina en la posición de aminoácidos 840) en el dominio HNH de Cas9 de *S. pyogenes* puede convertir a Cas9 en una nickasa. Otros ejemplos de mutaciones que convierten Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Ver, por ejemplo, Sapranaukas y otros, (2011) Nucleic Acids Research 39:9275-9282 y el documento WO 2013/141680. Dichas mutaciones pueden generarse utilizando métodos bien conocidos, como la mutagénesis dirigida al sitio, la mutagénesis mediada por PCR o la síntesis génica total. Pueden encontrarse ejemplos de otras mutaciones que crean nickasas, por ejemplo, en el documento WO/2013/176772A1 y el documento WO/2013/142578A1.

Las proteínas Cas pueden ser también proteínas de fusión. Por ejemplo, una proteína Cas puede fusionarse con un dominio de escisión, un dominio de modificación epigenética, un dominio de activación transcripcional o un dominio represor transcripcional. Ver el documento WO 2014/089290. Las proteínas Cas pueden fusionarse además con un polipéptido heterólogo que proporciona estabilidad aumentada o disminuida. El dominio fusionado o el polipéptido heterólogo pueden localizarse en el extremo N, el extremo C o internamente dentro de la proteína Cas.

Un ejemplo de una proteína de fusión Cas es una proteína Cas fusionada a un polipéptido heterólogo que proporciona la ubicación subcelular. Dichas secuencias pueden incluir, por ejemplo, una señal de ubicación nuclear (NLS) tal como el SV40 NLS para direccionarla al núcleo, una señal de ubicación mitocondrial para direccionarla a las mitocondrias, una señal de retención de ER y similares. Ver, por ejemplo, Lange y otros, (2007) J. Biol. Chem. 282:5101-5105. Una proteína Cas puede comprender, por ejemplo, una o más señales de ubicación nuclear (por ejemplo, dos señales de ubicación nuclear). Dichas señales de ubicación subcelular pueden localizarse en el N-terminal, el C-terminal 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 pueden comprender además 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 un secuencia de péptido de poliarginina. Ver, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular puede localizarse en el N-terminal, el C-terminal o en cualquier lugar dentro de la proteína Cas.

Las proteínas Cas pueden comprender además un polipéptido heterólogo para facilitar el seguimiento o la purificación, tales como una proteína fluorescente, una etiqueta de purificación o una etiqueta de epítipo. Los ejemplos de proteínas fluorescentes incluyen las proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Azami Green monomérica, CopGFP, AceGFP, ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrina, Venus, YPet, PhiYFP, ZsYellow1), proteínas fluorescentes azules (por ejemplo, eBFP, eBFP2, Azurita, mKalamal, GFPuv, zafiro, zafiro T), proteínas ciano fluorescentes (por ejemplo, eCFP, cerúleo, CyPet, AmCyan1, ciano Midoriishi), proteínas fluorescentes rojas (mKate, mKate2, mPlum, monómero de DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, monómero de DsRed, HcRed-Tándem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred) mKO, Kusabira Naranja, Kusabira Naranja monomérica, 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 carboxilo biotina (BCCP) y calmodulina.

Las proteínas Cas pueden proporcionarse en cualquier forma. Por ejemplo, puede proporcionarse una proteína Cas en forma de una proteína, tal como una proteína Cas complejada con un ARNg. Alternativamente, puede proporcionarse una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero

(ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas puede ser codón optimizado para la traducción eficiente en proteína en una célula u organismo en particular. Cuando un ácido nucleico que codifica la proteína Cas se introduce en la célula, la proteína Cas puede expresarse de manera transitoria, condicional o constitutiva en la célula.

Los ácidos nucleicos que codifican las proteínas Cas pueden integrarse de manera estable en el genoma de la célula y enlazarse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican las proteínas Cas pueden enlazarse operativamente a un promotor en un constructo de expresión. Los constructos de expresión incluyen cualquiera de los constructos de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. 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 está separado del vector que comprende el ADN que codifica el ARNg. Los promotores que pueden utilizarse en un constructo de expresión incluyen, por ejemplo, promotores activos en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Los ejemplos de otros promotores se describen en otra parte de la presente descripción.

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 objetivo. 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 molécula de ARN simple (polinucleótido de ARN simple), que puede denominarse además un "ARNg de molécula simple", un "ARN guía simple" o un "ARNgs." Ver, por ejemplo, los documentos WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622A2, WO/2014/099750A2, WO/2013142578A1 y WO 2014/131833A1. Los términos "ARN guía" y "ARNg" son inclusivos, que incluyen tanto los ARNg de doble molécula como los ARNg de molécula simple.

Un ARNg de dos moléculas ilustrativo comprende una molécula similar a un ARNcr ("ARN de CRISPR" o "ARN direccionador" o "ARNcr" o "repetición de ARNcr") y una molécula correspondiente similar al ARNtracr ("ARN de CRISPR de acción trans" o "ARN activador" o "ARNtracr" o "andamio"). Un ARNcr comprende tanto el segmento de direccionamiento al ADN (simple cadena) del ARNg como un tramo de nucleótidos que forma la mitad del dúplex de ARNbc del segmento de unión a proteínas del ARNg.

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 son complementarios y se hibridan 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.

El ARNcr y el ARNtracr correspondiente se hibridan para formar un ARNg. El ARNcr proporciona adicionalmente el segmento de direccionamiento al ADN de simple cadena que se hibrida con una secuencia objetivo. Si se usa para la modificación dentro de una célula, la secuencia exacta de una molécula de ARNcr o ARNtracr dada puede diseñarse para que sea específica de la especie en la que se usarán las moléculas de ARN. Ver, por ejemplo, Mali y otros, (2013) Science 339:823-826; Jinek y otros. (2012) Science 337:816-821; Hwang y otros, (2013) Nat. Biotechnol. 31:227-229; Jiang y otros. (2013) Nat. Biotechnol. 31:233-239; y Cong y otros, (2013) Science 339:819-823.

El segmento de direccionamiento al ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo. El segmento de direccionamiento al ADN de un ARNg interactúa con un ADN objetivo de una manera específica de secuencia a través de 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 objetivo con el que interactuarán el ARNg y el ADN objetivo. El segmento de direccionamiento al ADN de un ARNg sujeto puede modificarse para hibridarse con cualquier secuencia deseada dentro de un ADN objetivo. Los ARNcr de origen natural difieren en dependencia del sistema Cas9 y el organismo, pero frecuentemente contienen un segmento dirigido de entre 21 y 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (ver, por ejemplo, el documento WO2014/131833). En el caso de *S. pyogenes*, los DR tienen 36 nucleótidos de largo y el segmento de orientación tiene 30 nucleótidos de largo. El DR localizado en 3' es complementario y se 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 hasta aproximadamente 40 nt, desde aproximadamente 20 nt hasta aproximadamente 45 nt, desde aproximadamente 20 nt hasta aproximadamente 50 nt, desde aproximadamente 20 nt hasta aproximadamente 60 nt, desde aproximadamente 20 nt hasta aproximadamente 70 nt, desde aproximadamente 20 nt hasta 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 complementario a una secuencia de nucleótidos (secuencia objetivo) del ADN objetivo puede tener una longitud de al menos aproximadamente 12 nt. Por ejemplo, la secuencia dirigida al ADN (es decir, la secuencia dentro del segmento de direccionamiento al ADN que es complementaria a una secuencia objetivo dentro del ADN objetivo) 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, el segmento de direccionamiento al ADN puede tener una longitud de aproximadamente 12 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 hasta aproximadamente 50 nt, desde aproximadamente 19 nt hasta aproximadamente 60 nt, desde aproximadamente 20 nt hasta aproximadamente 25 nt, desde aproximadamente 20 nt hasta aproximadamente 30 nt, desde aproximadamente 20 nt hasta aproximadamente 35 nt, desde aproximadamente 20 nt hasta 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 dirigida al ADN puede tener una longitud de aproximadamente 20 nt.

Los ARNtracr pueden estar en cualquier forma (por ejemplo, los ARNtracr de longitud completa o ARNtracr parciales activos) y longitudes que varían. Pueden incluir transcriptos primarios o formas procesadas. Por ejemplo, los ARNtracr (como parte de un ARN de una sola guía o como una molécula separada como parte de un ARNg de dos moléculas) pueden comprender o consistir en la totalidad o una porción de una secuencia de ARNtracr de tipo silvestre (por ejemplo, aproximadamente o más de aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85 o más nucleótidos de una secuencia de ARNtracr de tipo silvestre). Los ejemplos de secuencias de ARNtracr de tipo silvestre de *S. pyogenes* incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos y 65 nucleótidos. Ver, por ejemplo, Deltcheva y otros, (2011) *Naturaleza* 471:602-607; el documento WO 2014/093661. Los ejemplos de ARNtracr dentro de los ARN guía simples (ARNgs) incluyen los segmentos de ARNtracr que se encuentran dentro de las versiones de ARNgs +48, +54, +67 y +85, donde "+n" indica que hasta el nucleótido +n del ARNtracr de tipo silvestre está incluido en el ARNgs. Ver el documento US 8,697,359.

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

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 sujeto interactúa con una proteína Cas, y el ARNg dirige la unión de la proteína Cas a una secuencia de nucleótidos específica dentro del ADN objetivo a través del segmento de direccionamiento al ADN.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada; orientación subcelular; seguimiento con un marcador fluorescente; un sitio de unión para una proteína o complejo de proteínas; y similares). Los ejemplos de tales modificaciones incluyen, por ejemplo, una caperuza en 5' (por ejemplo, una caperuza de 7-metilguanilato (m7G)); una cola poliadenilada 3' (es decir, una cola

5 poli(A) 3'); una secuencia de riboswitch (por ejemplo, para permitir estabilidad regulada y/o 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 (por ejemplo, núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona seguimiento (por ejemplo, conjugación directa a una molécula fluorescente, conjugación a una porción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.); una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, que incluyen activadores transcripcionales, represores transcripcionales, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares); y sus combinaciones.

10 Los ARN guía pueden proporcionarse de cualquier forma. Por ejemplo, el ARNg puede proporcionarse en forma de ARN, ya sea como dos moléculas (un ARNcr y ARNtracr separado) o como una molécula (ARNgs), y opcionalmente en forma de un complejo con una proteína Cas. El ARNg puede proporcionarse además en la forma de ADN que codifica el ARNg. El ADN que codifica el ARNg puede codificar una sola molécula de ARN (ARNgs) o moléculas de ARN separadas (por ejemplo, CrRNA y ARNtracr separados). En el último caso, el ADN que codifica el ARNg puede proporcionarse como moléculas de ADN separadas que codifican el ARNcr y el ARNtracr, respectivamente. Alternativamente, el ADN que codifica el ARNg puede proporcionarse como una molécula de ADN.

20 Cuando un ADN que codifica un ARNg se introduce en una célula, el ARNg puede expresarse de manera transitoria, condicional o constitutiva en la célula. Los ARNg que codifican las proteínas Cas pueden integrarse de manera estable en el genoma de la célula y enlazarse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg pueden enlazarse operativamente a un promotor en un constructo 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 está separado del vector que comprende el ácido nucleico que codifica la proteína Cas. Los promotores que pueden utilizarse en tales constructos de expresión incluyen promotores activos, por ejemplo, en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Tales 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.

30 Alternativamente, los ARNg pueden prepararse por varios otros métodos. Por ejemplo, los ARNg pueden prepararse mediante transcripción *in vitro* con el uso, por ejemplo, de ARN polimerasa T7 (ver, por ejemplo, el documento WO 2014/089290 y el documento WO 2014/065596). Los ARN guía pueden ser además una molécula producida sintéticamente preparada por síntesis química.

35 Una secuencia objetivo para un sistema CRISPR/Cas incluye las secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento de ADN de un ARNg, siempre que existan condiciones suficientes para la unión. Por ejemplo, las secuencias objetivo incluyen secuencias para las cuales un ARN guía se diseña para tener complementariedad, donde la hibridación entre una secuencia objetivo y una secuencia dirigida al ADN promueve la formación de un complejo CRISPR. No se requiere necesariamente una completa complementariedad, siempre que haya suficiente complementariedad para causar hibridación y promover la formación de un complejo CRISPR. Las secuencias objetivo incluyen también los sitios de escisión para las proteínas Cas, que se describen con más detalle más abajo. Una secuencia objetivo 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, como una mitocondria o cloroplasto.

45 La secuencia objetivo dentro de un ADN objetivo puede dirigirse por (es decir, estar unida por o hibridarse con, 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 a ADN/ARN adecuadas (por ejemplo, condiciones en un sistema libre de células) (ver, por ejemplo, *Molecular Cloning: A Laboratory Manual*, 3ra Ed. (Sambrook y otros, Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria y se hibrida con la proteína Cas o el ARNg puede denominarse la "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto, no es complementaria a la proteína Cas o ARNg) puede denominarse "cadena no complementaria" o "cadena de molde."

55 La proteína Cas puede escindir el ácido nucleico en un sitio dentro o fuera de una secuencia de ácido nucleico presente en un ADN objetivo al que se unirá un segmento de ADN de un ARNg. El "sitio de escisión" incluye la posición de un ácido nucleico en donde una proteína Cas produce una ruptura de cadena simple o una ruptura de doble cadena. Por ejemplo, la formación de un complejo CRISPR (que comprende un ARNg hibridado con una secuencia objetivo y complejo con una proteína Cas) puede provocar la escisión de una o ambas cadenas en o cerca de (por ejemplo, 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 objetivo al que se unirá un segmento de ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico presente en un ADN objetivo al que se unirá un segmento de ADN de un ARNg, el sitio de escisión todavía se considera dentro de la "secuencia objetivo." 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 pueden producirse, por ejemplo, con el uso de dos proteínas Cas que producen una ruptura de

cadena simple en diferentes sitios de escisión en cadenas diferentes, produciendo así, una ruptura de doble cadena. Por ejemplo, una primera nickasa puede crear una ruptura de una simple cadena en la primera cadena de ADN de doble cadena (ADNbc), mientras que una segunda nickasa puede crear una ruptura de una sola cadena en la segunda cadena de ADNbc, de modo que se crean secuencias salientes. En algunos casos, la secuencia objetivo de la nickasa en la primera cadena está separada de la secuencia objetivo 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 de sitio del ADN objetivo por Cas9 puede producirse en ubicaciones determinadas tanto por (i) complementariedad de apareamiento de bases entre el ARNg y el ADN objetivo como por (ii) un motivo corto, denominado el motivo adyacente al protoespaciador (PAM), en el ADN objetivo. El PAM puede flanquear la secuencia objetivo. Opcionalmente, la secuencia objetivo 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) corriente arriba o corriente abajo de la secuencia PAM. En algunos casos (por ejemplo, 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 está inmediatamente 3' de la secuencia objetivo de la cadena no complementaria del ADN objetivo. 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 objetivo de la cadena complementaria del ADN objetivo. En algunos casos, N<sub>1</sub> y N<sub>2</sub> puede ser complementarios y el par de bases N<sub>1</sub>-N<sub>2</sub> pueden ser cualquier par de bases (por ejemplo, N<sub>1</sub>=C y N<sub>2</sub>=G; N<sub>1</sub>=G y N<sub>2</sub>=C; N<sub>1</sub>=A y N<sub>2</sub>=T; o N<sub>1</sub>=T, y N<sub>2</sub>=A).

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

La secuencia objetivo puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia objetivo puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora o ADN redundante) o puede incluir ambos.

Se proporcionan además las variantes y fragmentos activos de los agentes tipo nucleasa (es decir, un agente tipo nucleasa diseñado por ingeniería genética). Dichas 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 tipo nucleasa nativo, en donde las variantes activas retienen la capacidad de cortar en un sitio de reconocimiento deseado y, por lo tanto, retienen la actividad de inducir mella o ruptura de doble cadena. Por ejemplo, cualquiera de los agentes tipo nucleasa descritos en la presente descripción puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir una ruptura o mella de doble cadena en un sitio de reconocimiento que no fue reconocido por el agente tipo nucleasa nativo. Por lo tanto, en algunas modalidades, la nucleasa diseñada por ingeniería genética tiene una especificidad para inducir una ruptura o mella de doble cadena en un sitio de reconocimiento que es diferente del sitio de reconocimiento del agente tipo nucleasa nativo correspondiente. Los ensayos de actividad inductora de mella o ruptura de doble cadena son conocidos y generalmente miden la actividad general y la especificidad de la endonucleasa en sustratos de ADN que contienen el sitio de reconocimiento.

El agente tipo nucleasa puede introducirse en la célula pluripotente por cualquier medio conocido en la técnica. El polipéptido que codifica el agente tipo nucleasa puede introducirse directamente en la célula. Alternativamente, un polinucleótido que codifica el agente tipo nucleasa puede introducirse en la célula. Cuando se introduce en la célula un polinucleótido que codifica el agente tipo nucleasa, el agente tipo nucleasa puede expresarse de forma transitoria, condicional o constitutiva dentro de la célula. Por lo tanto, el polinucleótido que codifica el agente tipo nucleasa puede estar contenido en un casete de expresión y estar operativamente enlazado a un promotor condicional, un promotor inducible, un promotor constitutivo o un promotor específico de tejido. Alternativamente, el agente tipo nucleasa se introduce en la célula como un ARNm que codifica un agente tipo nucleasa.

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

Cuando el agente tipo nucleasa se proporciona a la célula mediante la introducción de un polinucleótido que codifica el agente tipo nucleasa, dicho polinucleótido que codifica un agente tipo nucleasa puede modificarse para sustituir codones que tienen una mayor frecuencia de uso en la célula de interés, en comparación con la secuencia polinucleotídica de origen natural que codifica el agente tipo nucleasa. Por ejemplo, un ácido nucleico que codifica un agente tipo nucleasa puede modificarse para sustituir codones que tienen una mayor frecuencia de uso en una célula procariota o eucariota

dada, que incluye una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, un 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 polinucleótido de origen natural.

5 Los diversos métodos expuestos anteriormente pueden repetirse secuencialmente para permitir la integración dirigida de cualquier cantidad de insertos de ácido nucleico en un locus genómico a transformar 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 objetivo en un cromosoma. En modalidades particulares, tales métodos de mosaico secuencial permiten la reconstrucción de grandes regiones genómicas a partir 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 a transformar en un cromosoma. En tales casos, la transferencia y reconstrucción de las regiones genómicas que incluyen las regiones codificantes y no codificantes permiten preservar la complejidad de una región dada al retener, al menos en parte, las regiones codificantes, las regiones no codificantes y variaciones del número de copias encontradas dentro de la región genómica nativa. Así, los diversos métodos proporcionan, por ejemplo, los métodos para generar regiones genómicas "heterólogas" o "exógenas" dentro de una célula.

#### IV. Marcadores de selección

20 Los diversos métodos y composiciones proporcionados en la presente descripción pueden emplear los agentes tipo nucleasa y sus sitios de reconocimiento correspondientes en combinación con los marcadores de selección. Como se discute en la presente descripción, la posición del sitio de reconocimiento en el polinucleótido que codifica el marcador de selección permite un método eficiente por el cual identificar los eventos de integración en el locus objetivo. Además, en la presente descripción se proporcionan varios métodos en los que se emplean marcadores de selección alternativos que tienen el sitio de reconocimiento de nucleasas para mejorar la eficiencia y la eficacia a través de los cuales se integran múltiples polinucleótidos de interés dentro de un locus dirigido dado.

Pueden utilizarse varios marcadores de selección en los métodos y composiciones descritos en la presente descripción. Tales marcadores de selección pueden, por ejemplo, impartir resistencia a un antibiótico tales como G418, higromicina, blasticidina, neomicina, o puromicina. Tales 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 blasticidina S desaminasa (*bsr<sup>r</sup>*). Aún en otras modalidades, el marcador de selección se enlaza operativamente a un promotor inducible y la expresión del marcador de selección es tóxica para la célula. Los ejemplos no limitativos de tales marcadores de selección incluyen xantina/guanina fosforibosil transferasa (*gpt*), hipoxantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus herpes simple (HSV-TK).

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

40 El polinucleótido que codifica los marcadores de selección está operativamente enlazado a un promotor activo en la célula. Dichos casetes de expresión y sus diversos componentes reguladores se analizan en más detalle en otra parte de la presente descripción.

#### V. Promotores

45 Varias secuencias de ácido nucleico descritas en la presente descripción pueden enlazarse operativamente a promotores. Tales promotores que pueden ser activos, por ejemplo, en una célula pluripotente, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Un promotor puede ser, por ejemplo, un promotor constitutivamente activo, un promotor condicional, un promotor inducible, un promotor temporalmente restringido (por ejemplo, un promotor regulado por el desarrollo) o un promotor espacialmente restringido (por ejemplo, un promotor específico de una célula o específico de tejido). Pueden encontrarse ejemplos de promotores, por ejemplo, en el documento WO 2013/176772.

Los ejemplos de promotores inducibles incluyen, por ejemplo, los promotores regulados químicamente y promotores regulados físicamente. Los promotores regulados químicamente incluyen, por ejemplo, promotores regulados por alcohol (por ejemplo, un promotor génico de la alcohol deshidrogenasa (*alcA*)), promotores regulados por tetraciclina (por ejemplo, un promotor sensible a la tetraciclina, una secuencia de operador de tetraciclina (*tetO*), un promotor *tet-On*, o un promotor *tet-Off*), promotores regulados por esteroides (por ejemplo, un receptor de glucocorticoides de rata, un promotor de un receptor de estrógenos, o un promotor de un receptor de ecdisona), o promotores regulados por metales (por ejemplo, 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 represible por luz).

Los promotores específicos de tejido pueden ser, por ejemplo, promotores específicos de neurona, 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 inmunes (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.

Puede seleccionarse además un promotor en función del 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 humana adulta, una célula progenitora humana con desarrollo restringido, 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

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

Además se proporcionan polinucleótidos recombinantes que comprenden los diversos componentes del sistema de transformación. Los términos "polinucleótido recombinante" y "constructo de ADN recombinante" se usan indistintamente en la presente descripción. Un constructo recombinante comprende una combinación artificial o heteróloga de secuencias de ácido nucleico, por ejemplo, secuencias reguladoras y de codificación que no se encuentran juntas en la naturaleza. En otras modalidades, un constructo 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 manera diferente a la encontrada en la naturaleza. Tal constructo puede utilizarse solo o puede utilizarse junto con un vector. Si se usa un vector, entonces la elección del vector depende del método que se usa para transformar las células huésped como es bien conocido por los expertos en la técnica. Por ejemplo, puede utilizarse un vector plasmídico. En la presente descripción se proporcionan elementos genéticos necesarios para transformar, seleccionar y propagar con éxito las células huésped y que comprenden cualquiera de los fragmentos de ácido nucleico aislados. La detección 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.

En modalidades específicas, uno o más de los componentes del sistema de transformación descrito en la presente pueden proporcionarse 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' enlazadas operativamente a un polinucleótido proporcionado en la presente. Cuando se utiliza para referirse a la unión de dos regiones de codificación de proteínas, operativamente enlazadas significa que las regiones de codificación están en el mismo marco de lectura. En otro caso, una secuencia de ácido nucleico que codifica una proteína puede estar operativamente enlazada a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) para retener la regulación transcripcional adecuada.

El casete puede contener adicionalmente al menos un polinucleótido adicional de interés para ser introducido conjuntamente en el organismo. Alternativamente, el polinucleótido adicional de interés puede proporcionarse en casetes de expresión múltiple. Tal 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 está bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes marcadores de selección.

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 la presente descripción, y una región de terminación transcripcional y traduccional (es decir, región de terminación) funcional en la célula de mamífero o una célula huésped de interés. Las regiones reguladoras (es decir, los promotores, las regiones reguladoras de la transcripción y las regiones de terminación de la traducción) y/o un polinucleótido proporcionado en la presente descripción pueden ser nativos/análogos a la célula huésped o entre sí. Alternativamente, las regiones reguladoras y/o un polinucleótido proporcionado en la presente descripción pueden ser heterólogas a la célula huésped o entre sí. Por ejemplo, un promotor enlazado operativamente a un polinucleótido heterólogo es de una especie diferente de la especie de la que se deriva el polinucleótido o, si es de la misma especie/análoga, uno o ambos están sustancialmente modificados de su forma original y/o locus, o el promotor no es el promotor nativo del polinucleótido operativamente enlazado. Alternativamente, las regiones reguladoras y/o un polinucleótido recombinante proporcionado en la presente descripción pueden ser completamente sintéticos.

La región de terminación puede ser nativa con la región de inicio de la transcripción, puede ser nativa con el polinucleótido recombinante enlazado operativamente, puede ser nativa con la célula huésped, o puede derivarse de otra fuente (es decir, foránea o heteróloga) al promotor, el recombinante polinucleótido, la célula huésped, o cualquiera de sus combinaciones.

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 u otras manipulaciones pueden estar implicadas para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción, o similares. Para este propósito, pueden estar involucradas mutagénesis in vitro, reparación de cebadores, restricción, hibridación, resustituciones, por ejemplo, transiciones y transversiones.

5 Pueden usarse varios promotores en los casetes de expresión proporcionados en la presente descripción. Los promotores pueden seleccionarse 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 momento, la ubicación y/o el nivel de expresión del polinucleótido de interés. Tales constructos de expresión también pueden contener, si se desea, una  
10 región reguladora del promotor (por ejemplo, una que confiera expresión inducible, constitutiva, regulada ambientalmente o por el desarrollo, o específica/selectiva de células o tejidos), un sitio de iniciación de la transcripción, un sitio de unión a ribosoma, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación.

15 El casete de expresión que contiene los polinucleótidos proporcionados en la presente descripción puede comprender además 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.

20 Cuando sea apropiado, las secuencias empleadas en los métodos y composiciones (es decir, el polinucleótido de interés, el agente tipo nucleasa, etc.) pueden optimizarse para aumentar la expresión en la célula. Es decir, los genes pueden sintetizarse con el uso de codones preferidos en una célula de interés dada, que incluyen, por ejemplo, codones preferidos de mamíferos, codones preferidos de humanos, codones preferidos de roedores, codones preferidos de ratones, codones preferidos de ratas, etc. para la expresión mejorada.

25 En una modalidad, el agente tipo nucleasa se expresa a partir de un constructo de expresión que comprende una secuencia de ácido nucleico que codifica una nucleasa, y el ácido nucleico se enlaza operativamente a un promotor activo en la célula.

#### VII. Métodos para producir animales no humanos genéticamente modificados

30 Los animales no humanos modificados genéticamente pueden generarse mediante el empleo de los diversos métodos descritos en la presente descripción. En algunos casos, el método para producir un animal no humano genéticamente modificado comprende: (1) modificar el genoma de una célula pluripotente no humana con el uso de los métodos descritos en la presente; (2) seleccionar la célula pluripotente no humana genéticamente modificada; (3) introducir la célula pluripotente no humana genéticamente modificada en un embrión huésped no humano; y (4) implantar el embrión huésped  
35 no humano que comprende la célula pluripotente no humana genéticamente modificada en una madre sustituta no humana. Se genera una progenie de la célula pluripotente no humana genéticamente modificada. La célula donante puede introducirse en un embrión huésped en cualquier etapa, tal como la etapa de blastocito o la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). La progenie que es capaz de transmitir la modificación genética a través de la línea germinal se genera. La célula pluripotente puede ser, por ejemplo, una célula ES (por ejemplo, una célula ES de ratón, o una célula ES de rata) como se describe en otras partes en la presente descripción. Ver, por ejemplo, patente de Estados Unidos núm. 7.294.754.

45 Las técnicas de transferencia nuclear pueden utilizarse además para generar los animales de mamíferos no humanos. Brevemente, los métodos para la transferencia nuclear pueden incluir las etapas de: (1) enucleado de un oocito no humano o proporcionar un oocito enucleado no humano; (2) aislar o proporcionar una célula o núcleo donante no humano a combinarse con el oocito enucleado; (3) insertar la célula o el núcleo en el oocito 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) dejar que se desarrolle el embrión. En tales métodos, los oocitos se recuperan generalmente de animales fallecidos, aunque pueden aislarse también de oviductos y/u ovarios de animales no humanos vivos. Antes de la enucleación los oocitos pueden madurarse en una variedad de medios conocidos por los expertos en la técnica. La enucleación del oocito puede realizarse de una serie de maneras bien conocidas por los expertos en la técnica. La inserción de la célula o núcleo donante no humano en el oocito enucleado no humano para formar una célula reconstituida puede realizarse mediante microinyección de una célula donante no humana debajo de la zona pelúcida antes de la fusión. La fusión puede inducirse por la aplicación de un pulso eléctrico DC a través del plano de contacto/fusión (electrofusión), por exposición de las  
50 células no humanas a sustancias químicas promotoras de la fusión, tales como polietilenglicol, o por medio de un virus inactivado, tal como el virus de Sendai. Una célula reconstituida puede activarse por medios eléctricos y/o no eléctricos antes, durante, y/o después de la fusión del donante nuclear y el oocito receptor. Los métodos de activación incluyen pulsos eléctricos, un choque inducido químicamente, penetración por esperma, niveles crecientes de cationes divalentes en el oocito, y reducción de la fosforilación de proteínas celulares (como por medio de inhibidores de quinasas) en el oocito. Las células no humanas reconstituidas activadas, o embriones, pueden cultivarse en medio conocido por los expertos en la técnica y después transferirse al útero de un animal no humano. Ver, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1y patente de los Estados Unidos núm. 7,612,250.

65 Los métodos de esta descripción 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 producir el animal no humano de 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 después implantarse en una madre sustituta para producir un animal F0. Los animales que portan el locus genómico genéticamente modificado pueden identificarse mediante un ensayo de modificación del alelo (MOA) como se describe en la presente descripción.

Los diversos métodos proporcionados en la presente descripción permiten la generación de un animal no humano F0 genéticamente modificado en donde las células del animal F0 genéticamente modificado comprenden la modificación dirigida. Se reconoce que en dependencia 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 donantes en un embrión en etapa premórula a partir de un organismo correspondiente (por ejemplo, un embrión de ratón en una etapa de 8 células) a través, por ejemplo, del método VELOCIMOUSE® permite que un mayor porcentaje de la población celular del animal F0 comprenda las células que tienen la modificación genética dirigida. Por ejemplo, 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 no humano F0 puede comprender una población celular con la modificación dirigida. Adicionalmente, al menos una o más de las células germinales del animal F0 pueden tener la modificación dirigida.

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

En algunos casos, el animal F0 generado por los métodos y composiciones de esta descripción puede criarse como un animal de tipo silvestre para generar una generación F1 que sea heterocigota para la modificación dirigida. Los animales de la generación F1 pueden criarse entre sí para generar un animal F2 homocigoto para la modificación dirigida. La progenie F1 puede genotiparse con el uso de cebadores y/o sondas específicos para determinar si la modificación genética dirigida está presente.

#### VIII. Métodos para introducir ácidos nucleicos y proteínas en las células

En la presente descripción se proporcionan diversos 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 tipo nucleasa, sitios de escisión de nucleasa, insertos de ácido nucleico, vectores de transformación, loci genómicos objetivo y polinucleótidos de interés).

Los métodos proporcionados en la presente descripción pueden comprender la introducción en una célula de uno o más polinucleótidos o constructos de polipéptidos 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 gane acceso al interior de la célula. Los métodos proporcionados en la presente descripción no dependen de un método particular para introducir un ácido nucleico en la célula, solo que el ácido nucleico o la proteína gane acceso al interior de al menos una célula. Los métodos para introducir ácidos nucleicos y proteínas en varios tipos de células se conocen en la técnica e incluyen, por ejemplo, métodos de transfección estable, métodos de transfección transitoria, y métodos mediados por virus.

En algunos casos, las células empleadas en los métodos y composiciones tienen un constructo de ADN incorporado de manera estable en su genoma. "Establemente incorporado" o "establemente introducido" 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 pueda ser heredada por la progenie de esta. Puede utilizarse 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 células pueden variar. Los métodos de transfección incluyen métodos de transfección a base de químicos con el uso de liposomas; nanopartículas; fosfato de calcio (Graham y otros, (1973) Virology 52 (2): 456-67, Bacchetti y otros, (1977) Proc Natl Acad Sci USA 74 (4): 1590-4, y Krieglner, M (1991). Transfer and Expression: A Laboratory Manual. New York: W. H. Freeman y Compañía. páginas 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilenimina. Los métodos no químicos incluyen electroporación, sonoporación; y transfección óptica. La transfección basada en partículas incluye el uso de una transfección con pistola de genes o asistida por imanes (Bertram (2006) Current Pharmaceutical Biotechnology 7, 277-28). Los métodos virales pueden utilizarse además para la transfección.

En algunos casos, la introducción de ácidos nucleicos en una célula también puede medirse por electroporación, por inyección intracitoplasmática, por infección viral, por adenovirus, por virus adenoasociado, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos, o por Nucleofection™.

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

Tanto los agentes tipo nucleasa como los vectores de transformación (por ejemplo, los LTVEC) pueden introducirse simultáneamente cuando se introducen en la célula. Alternativamente, el agente tipo nucleasa puede introducirse por separado de los vectores de transformación. Por ejemplo, el agente tipo nucleasa puede introducirse antes de la introducción de los vectores de transformación, o puede introducirse después de la introducción de los vectores de transformación. Cuando se introducen dos o más LTVEC en la célula, pueden introducirse simultáneamente, o alternativamente, pueden introducirse por separado.

## IX. Células y Animales

Diversas composiciones y métodos proporcionados en la presente descripción emplean células, tales como células de un animal. Dichas células pueden ser células no humanas que pueden ser de un animal no humano. Dichas células pueden ser células eucariotas, que incluyen, por ejemplo, células fúngicas (por ejemplo, levadura), células vegetales, células animales, células de mamíferos y células humanas. 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 no humana, 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 no humanas incluyen células indiferenciadas que pueden dar lugar a cualquier tipo de célula, y las células pluripotentes incluyen células indiferenciadas que poseen la capacidad de convertirse en más de un tipo celular diferenciado. Tales células totipotentes y/o pluripotentes no humanas 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 que son capaces de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión. Las células ES 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 pueden ser además células madre hematopoyéticas o células madre neuronales.

Una célula eucariota puede ser además 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 no diferenciada.

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

Las células eucariotas incluyen además las células inmortalizadas. Las células inmortalizadas incluyen células de un organismo multicelular que normalmente no proliferaría indefinidamente pero, debido a la mutación o la alteración, han evadido la senescencia celular normal y, en cambio, pueden seguir sufriendo división. Tales mutaciones o alteraciones pueden ocurrir naturalmente o ser intencionalmente inducidas. Los ejemplos de células inmortalizadas incluyen las células de ovario de hámster chino (CHO), células de riñón embrionario humano (por ejemplo, células HEK 293) y células de fibroblastos embrionarios de ratón (por ejemplo, células 3T3). Numerosos tipos de células inmortalizadas se conocen bien en la técnica.

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

El término "animal", en referencia a células, células totipotentes pluripotentes y/o no humanas, células ES, células donadoras y/o embriones del 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, hámsters, cobayas), 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. Se incluyen además los animales domésticos y los animales agrícolas. El término "animal no humano" excluye a los seres humanos.

5 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 (por ejemplo, 129S1/SV, 129S1/Svlm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6 (129/SvEvTac), 129S7, 129S8, 129T1 y 129T2. Ver, por ejemplo, Festing y otros, (1999) *Mammalian Genome* 10:836). Los ejemplos de cepas C57BL incluyen C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Ka<sub>l</sub>wN, C57BL/6, C57BL/6J, C57BL/6ByJ, 10 C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr, y C57BL/Ola. Las células pluripotentes y/o totipotentes de ratón pueden ser además de una mezcla de una cepa 129 mencionada anteriormente y una cepa C57BL/6 mencionada anteriormente (por ejemplo, 50 % de 129 y 50 % de C57BL/6). De manera similar, las denominadas pluripotentes y/o totipotentes de ratón pueden ser de una mezcla de las cepas 129 mencionadas anteriormente o de 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 VGF1 de ratón. Ver, por ejemplo, Auerbach y otros, (2000) *Biotechniques* 29, 1024-1028, 15 1030, 1032.

Una célula pluripotente y/o totipotente de rata puede ser de cualquier cepa de rata, que incluye, 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 20 Dawley (SD), o una cepa de rata Fischer tal como Fisher F344 o Fisher F6. Las células pluripotentes y/o totipotentes de rata pueden obtenerse, además, a partir de cepas derivadas de una mezcla de dos o más cepas mencionadas anteriormente. Por ejemplo, una célula pluripotente y/o totipotente de rata puede ser de una cepa DA o una cepa ACI. Una cepa de rata ACI se caracteriza por tener agutí negro, con patas y vientre blanco y un haplotipo *RT1<sup>av1</sup>*. Tales cepas están disponibles a partir de una diversidad de fuentes que incluyen los laboratorios Harlan. Un ejemplo de una línea celular ES de rata de una rata ACI es una línea celular ES ACI.G1 de rata. Una cepa de rata Dark Agouti (DA) se caracteriza por tener un pelaje de agutí y un haplotipo *RT1<sup>av1</sup>*. Tales ratas están disponibles a partir de una diversidad de fuentes que incluyen los laboratorios Charles River y Harlan. Los ejemplos de una línea celular ES de rata a partir de una rata DA son la línea celular ES de rata DA.2B y la línea celular ES de rata DA.2C. En algunos casos, las células pluripotentes y/o totipotentes de ratas son de una cepa de rata endogámica. Ver, por ejemplo, el documento de patente US 2014/0235933 A1, presentado el 20 de febrero de 2014 y el documento de patente US 2014/0310828 A1, presentado el 16 de abril de 2014. 25 30

Los ejemplos de células pluripotentes humanas incluyen células ES humanas, células madre adultas humanas, células progenitoras humanas con desarrollo restringido y células madre pluripotentes inducidas por humanos (iPS), tales como 35 las células iPS humanas con cebado y células iPS humanas vírgenes. Ver, por ejemplo, la solicitud de patente de los Estados Unidos núm. 14/515,503, presentada el 15 de octubre. Las células madre pluripotentes inducidas incluyen células madre pluripotentes que pueden derivarse directamente de una célula adulta diferenciada. Las células iPS humanas pueden generarse mediante la introducción de conjuntos específicos de factores de reprogramación en una célula que pueden incluir, por ejemplo, Oct3/4, factores de transcripción de la familia Sox (por ejemplo, Sox1, Sox2, Sox3, Sox15), 40 factores de transcripción de la familia Myc (por ejemplo, c-Myc, l-Myc, n-Myc), factores de transcripción de la familia tipo Krüppel (KLF) (por ejemplo, KLF1, KLF2, KLF4, KLF5) y/o factores de transcripción relacionados, como NANOG, LIN28 y/o Glis1. Las células iPS humanas además pueden generarse, por ejemplo, mediante el uso de los miARN, moléculas pequeñas que imitan las acciones de 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 además se caracterizan por su capacidad de propagarse indefinidamente bajo condiciones de cultivo *in vitro* adecuadas. Ver, por ejemplo, Takahashi y Yamanaka (2006) *Cell* 126:663-676. Las células ES humanas con cebado y las células iPS humanas con cebado incluyen las células que expresan características similares a las de las células epiblasticas de postimplantación y están comprometidas en la especificación y diferenciación del linaje. Las células ES humanas vírgenes y las células iPS 50 humanas vírgenes incluyen las células que expresan características similares a las de las células ES de la masa celular interna de un embrión de preimplantación y no están comprometidas en la especificación del linaje. Ver, por ejemplo, Nichols y Smith (2009) *Cell Stem Cell* 4:487-492.

Las células que se han implantado en un embrión del huésped 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. De manera similar, la madre sustituta puede ser de la misma cepa que la célula pluripotente y/o totipotente modificada genéticamente y/o el embrión del huésped, o la madre sustituta puede ser de una cepa diferente que la célula pluripotente y/o totipotente genéticamente modificada y/o o el embrión del huésped. 55

Puede emplearse una variedad de embriones de huésped no humanos en los métodos y composiciones descritos en la presente. Por ejemplo, las células pluripotentes y/o totipotentes que tienen la modificación genética dirigida pueden introducirse en un embrión en etapa premórula (por ejemplo, un embrión en etapa de 8 células) de un organismo correspondiente. Ver, 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 donantes pueden implantarse en un embrión huésped en la 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 puede además ser un blastocisto o puede ser un embrión preblastocisto, un embrión en etapa premórula, un embrión en etapa 60 65

mórula, un embrión en etapa mórula no compactada o un embrión en etapa mórula compactada. Cuando se emplea un embrión de ratón, la etapa de embrión del 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 puede seleccionarse entre TS1, TS2, TS3 y TS4. En algunos casos, el embrión del huésped no humano comprende una zona pelúcida, y la célula donante no humana es una célula ES que se introduce en el embrión del huésped no humano a través de un agujero en la zona pelúcida. En otros casos, el embrión del huésped no humano es un embrión sin la zona. En aún otros casos, el embrión de huésped no humano en la etapa de mórula está agregado.

#### 10 X. Métodos de identificación de células con un locus genómico objetivo modificado

Algunos de los métodos anteriores comprenden además identificar una célula que tiene un locus genómico objetivo modificado (por ejemplo, un genoma modificado). Varios métodos pueden utilizarse para identificar las células que tienen una modificación dirigida, tal como una delección o una inserción. Tales métodos pueden comprender la identificación de una célula que tiene la modificación dirigida en un locus objetivo. Puede hacerse un examen para identificar tales células con los loci genómicos modificados.

La etapa de tamizaje puede comprender un ensayo cuantitativo para evaluar la modificación del alelo (MOA) de un cromosoma parental. Por ejemplo, el ensayo cuantitativo puede llevarse a cabo mediante una PCR cuantitativa, como una PCR en tiempo real (qPCR). La PCR en tiempo real puede utilizar un conjunto de primeros cebadores que reconoce el locus objetivo 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.

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. Dichas células se emplean después en los diversos métodos y composiciones descritos en la presente.

Otros ejemplos de ensayos cuantitativos adecuados incluyen la hibridación in situ mediada por fluorescencia (FISH), la hibridación genómica comparativa, la amplificación de ADN isotérmica, la hibridación cuantitativa a una(s) sonda(s) inmovilizada(s), Invader Probes®, MMP assays®, TAQMAN® Molecular Beacon, o tecnología de sonda Eclipse™ (ver, por ejemplo, el documento US2005/0144655).

La etapa de tamizaje puede comprender, además, ensayos específicos de brazo, que son ensayos usados para distinguir entre las inserciones dirigidas correctas de un inserto de ácido nucleico en un locus genómico objetivo a partir de inserciones transgénicas aleatorias del inserto de ácido nucleico en ubicaciones genómicas fuera del locus genómico objetivo se utilizan además para detectar el ensamblado correcto de dos o más LTVEC superpuestos en un único constructo. Los ensayos convencionales para el tamizaje de modificaciones dirigidas, tales como PCR de largo alcance, o transferencia de tipo Southern, vinculan el vector de transformación insertado al locus dirigido. Debido a sus tamaños grandes de brazo de homología, sin embargo, los LTVEC no permiten el tamizaje mediante tales ensayos convencionales. Para seleccionar la transformación por LTVEC, pueden usarse ensayos de modificación de alelos (MOA) que incluyen ensayos de pérdida de alelos (LOA) y de ganancia de alelo (GOA) (ver, por ejemplo, el documento US 2014/0178879 y Friendewey y otros, (2010) Methods Enzymol. 476:295-307). El ensayo de pérdida de alelos (LOA) invierte la lógica de tamizaje convencional y cuantifica el número de copias del locus nativo al que se dirigió la mutación. En un clon de células correctamente dirigidas, el ensayo LOA detecta uno de los dos alelos nativos (para genes que no están en el cromosoma X o Y), el otro alelo se ve interrumpido por la modificación dirigida. El mismo principio puede aplicarse en inverso como un ensayo de ganancia de alelo (GOA) para cuantificar el número de copias del vector de transformación insertado. Por ejemplo, el uso combinado de ensayos GOA y LOA revelará que un clon heterocigótico correctamente dirigido ha perdido una copia del gen objetivo natural y ganado una copia del gen de resistencia a fármacos u otro marcador insertado.

A modo de ejemplo, puede usarse la reacción en cadena de la polimerasa cuantitativa (qPCR) como el método de cuantificación de alelos, pero cualquier método que pueda distinguir de manera confiable la diferencia entre cero, una y dos copias del gen objetivo o entre cero, una y dos copias del inserto de ácido nucleico puede utilizarse para desarrollar un ensayo MOA. Por ejemplo, TAQMAN® pueden usarse para cuantificar el número de copias de un molde de ADN en una muestra de ADN genómico, especialmente mediante la comparación con un gen de referencia (ver, por ejemplo, el documento US 6,596,541). El gen de referencia se cuantifica en el mismo ADN genómico que el(los) gen(es) o locus(oci) objetivo(s). Por lo tanto, se realizan dos amplificaciones TAQMAN® (cada una con su sonda respectiva). 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(de los) gen(es) o locus(oci) dirigido(s) que se reemplaza por la transformación exitosa (es decir, un ensayo LOA). El Ct es una cantidad que refleja la cantidad de ADN de inicio para cada una de las sondas para el TAQMAN® es decir, una secuencia menos abundante requiere más ciclos de PCR para alcanzar el ciclo umbral. La disminución a la mitad del número de copias de la secuencia molde para una reacción del TAQMAN® dará como resultado un aumento de aproximadamente una unidad de Ct. Las reacciones del TAQMAN® en las células donde un alelo del(de los) gen(es) o locus(oci) objetivo(s) se han reemplazado por recombinación homóloga darán como resultado un aumento de un Ct para la reacción TAQMAN® objetivo sin un aumento en el Ct para el gen de referencia cuando se compara con el ADN de las células no dirigidas. Para un ensayo GOA, otra sonda TAQMAN® puede utilizarse para determinar el Ct del inserto de ácido nucleico que está reemplazando el(los) gen(s) o locus(oci) dirigido(s) mediante la transformación exitosa.

Puede ser un útil aumentar los ensayos LOA y GOA estándar para verificar transformación correcta por los LTVEC. Por ejemplo, los ensayos LOA y GOA solos no pueden distinguir los clones de células correctamente transformadas de los clones en los que una delección inducida por Cas del locus genómico objetivo coincide con la integración aleatoria de un LTVEC en cualquier otra parte del genoma. Debido a que la presión de transformación en la célula transformada se basa en el casete de selección, la integración transgénica aleatoria del LTVEC en cualquier otra parte del genoma generalmente incluirá el casete de selección y las regiones adyacentes del LTVEC pero pueden excluir las regiones más distales del LTVEC. Por ejemplo, si una porción 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 integrarán 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 integrarán transgénicamente con el casete de selección. A modo de ejemplo, si se usan ensayos LOA y GOA para evaluar la integración dirigida del LTVEC, y el ensayo GOA utiliza sondas contra el casete de selección o cualquier otra región única (sin brazo) del LTVEC, una delección heterocigota en el locus genómico objetivo combinado con una integración transgénica aleatoria del LTVEC dará la misma lectura que una integración dirigida heterocigota del LTVEC en el locus genómico objetivo. Para verificar la transformación correcta por el LTVEC, los ensayos específicos de brazos pueden utilizarse junto con los ensayos LOA y/o GOA.

Los ensayos específicos de los brazos determinan los números de copias de un molde de ADN en los brazos de homología de LTVEC. Dichos brazos de homología pueden incluir un brazo de homología de un LTVEC que no se superpone con otro LTVEC pero corresponde con una secuencia objetivo en la célula (por ejemplo, el brazo de homología que se superpone con la secuencia objetivo genómica en una célula de ratón (mArm)). Dichos brazos de homología pueden incluir además un brazo de homología superpuesto presente en dos LTVEC superpuestos (por ejemplo, la secuencia humana superpuesta en un brazo de homología 3' de un cebador LTVEC y un brazo de homología 5' de un segundo LTVEC (hArm)). Para los experimentos en los que se introducen múltiples LTVEC superpuestos en una célula, el tamizaje generalmente comprende los ensayos LOA, los ensayos GOA para todas las secuencias insertadas únicas y los ensayos específicos de brazo para todas las regiones de homología (es decir, entre el LTVEC y la secuencia objetivo 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 objetivo 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 la secuencia objetivo 5' de ratón), 1 copia de hArm1 (secuencia de superposición entre LTVEC 1 y 2), 1 copia de hArm2 (secuencia de superposición entre LTVEC 2 y 3) y 2 copias de mArm 3' (brazo de homología superpuesto con la secuencia objetivo 3' de ratón). En el ejemplo anterior, los números de copias de mArm mayores que dos generalmente indicarían una integración transgénica de LTVEC al azar fuera del locus genómico objetivo en lugar del locus genómico objetivo, lo cual es indeseable. Los clones dirigidos correctamente retendrían números de copia mArm de dos. Además, los números de copias de mArm de menos de dos en tales ensayos específicos de brazo generalmente indicarían delecciones mediadas por Cas grandes que se extienden más allá de la región dirigida para la delección, que también son indeseables. De manera similar, 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 un único constructo.

Si diferentes versiones de una secuencia se asocian con un número de registro en diferentes momentos, prevalecerá la versión asociada con el número de registro en la fecha de presentación efectiva de esta solicitud. La fecha de presentación efectiva significa la más temprana de la fecha de presentación real o fecha de presentación de una solicitud de prioridad que se refiere al número de registro si es aplicable. De manera similar, si diferentes versiones de una publicación, sitio web o lo similar, se publican en diferentes momentos, prevalecerá la versión más recientemente publicada con respecto a la fecha de presentación eficaz de la solicitud a menos que se indique de cualquier otra manera. Cualquier característica, etapa, elemento, modalidad o aspecto de la invención puede utilizarse en combinación con cualquier otro a menos que se indique específicamente de otra manera. 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 pueden practicarse ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

## EJEMPLOS

Los ejemplos siguientes se proponen para proporcionar a los expertos en la técnica una descripción y discusión completa de cómo realizar y usar los métodos y composiciones de la presente invención, y no pretenden limitar el alcance de lo que los inventores consideran como su invención ni pretenden representar que los experimentos más abajo son la totalidad o los únicos experimentos realizados. Se han realizado esfuerzos para asegurar la precisión con respecto a los números utilizados (por ejemplo, cantidades, temperaturas, etcétera) pero deben tenerse en cuenta algunos errores experimentales y desviaciones. A menos que se indique de cualquier otra manera, las partes son partes en peso, el peso molecular es peso molecular promedio, la temperatura es en grados Centígrados, y la presión es la atmosférica o cercana a esta.

Ejemplo 1: Transformación del locus TCR Alfa con dos LTVEC en combinación con las nucleasas de dedos de zinc

Se diseñó un sistema de doble transformación de manera que dos grandes vectores de transformación (LTVEC) modifiquen un locus genómico en una única etapa de transformación. Como se muestra en la Figura 1, una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de

selección de higromicina se dirigió por el sistema de doble transformación para generar las células ES que comprenden segmentos de genes variables I<sub>g</sub>k adicionales.

Este enfoque de doble transformación, se resume en la Figura 1, implica la doble transformación o la coelectroporación de dos vectores de transformación grandes diferentes (LTVEC) junto con las secuencias de nucleótidos que codifican una endonucleasa (por ejemplo, nucleasas de dedo de zinc) que crea una ruptura de doble cadena en o cerca del locus objetivo en las células ES.

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

Las nucleasas de dedos de zinc (ZFN) se diseñaron para reconocer y escindir una secuencia objetivo dentro del gen de resistencia a la higromicina para promover la recombinación homóloga de los dos LTVEC en el locus TCR A objetivo. Las células ES generadas en la Figura 1 (MAID 6548, heterocigoto para todos los segmentos de J<sub>κ</sub> humanos y cuatro segmentos de genes de V<sub>κ</sub> humanos funcionales) se electroporaron con los dos vectores de transformación grandes (MAID6600 y MAID1700-recortados) descritos anteriormente y dos plásmidos que expresan cada mitad de la ZFN (1/2), que se unen a las secuencias de reconocimiento en el gen de resistencia a la higromicina y catalizan una ruptura de doble cadena en el sitio objetivo (TGCGATCGCTGCGGCCGAtcttagCCAGACGAGCGGGTTCGG (SEQ ID NO: 2); con el sitio de escisión en letras minúsculas) (ver Tabla 1). Se diseñaron dos ZFN adicionales para la higromicina objetivo: ZFN (3/4) que se dirige al gen de higromicina en la secuencia de nucleótidos CGCTGCGGCCGATCTtagccaGACGAGCGGGTTCGG (SEQ ID NO: 3); y ZFN (5/6) que se dirige al gen de higromicina en la secuencia de nucleótidos AGCGTGCCGACCTGATGcagctcTCGGAGGGCGAAGAA (SEQ ID NO: 4) (ver Tabla 1).

Tabla 1: Sitios de unión y escisión de la nucleasa de dedos de zinc en la higromicina (cadena complementaria no mostrada)

Nucleasa de dedos de zinc	Medio dedo de zinc	Secuencia de unión (5'-3')	Sitio de escisión (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	AGCGTGCCGACCTGATG (SEQ ID NO: 9)	CAGCTC (SEQ ID NO: 13)
	Hyg-ZF6	TTCTTCGCCCTCCGA (SEQ ID NO: 10)	

Los dos vectores de transformación sometidos a electroporación conjunta se insertaron mediante la recombinación homóloga en el genoma de las células ES en la secuencia de higromicina, reemplazando la región que contiene y rodea el casete de selección de la higromicina. Las células ES resultantes contenían en el locus TCR A endógeno un dominio variable de la inmunoglobulina humana que comprende segmentos de genes humanos J<sub>κ</sub>1 a J<sub>κ</sub>5 y V<sub>κ</sub>4-1 a V<sub>κ</sub>2-30. La incorporación exitosa de los dos grandes vectores de transformación se confirmó con el uso de los ensayos TAQMAN® descritos anteriormente (Lie y Petropoulos, 1998. Curr. Opin. Biotechnology 9:43-48), con el uso de sondas e cebadores indicados en la Figura 1 y enumerado en la Tabla 2 más abajo (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 vector de transformación grande (MAID 1710); hArm2 = superposición de 20 kb de los primero (MAID 1710) y segundo (MAID 6600) vectores de transformación grandes, mArm = brazo de homología 5' de 15,5 kb del segundo vector de transformación (MAID 6600), control de ratón de tipo silvestre - secuencias presentes en el locus TCR A de ratón). Los ensayos de PCR en tiempo real que reconocen las secuencias en los brazos de homología de los LTVEC, denominados ensayos específicos del brazo, se utilizaron para verificar la orientación correcta del LTVEC en el genoma del ratón. La determinación del número de copias de estos ensayos específicos del brazo proporcionó un aclaramiento adicional para ayudar a distinguir los clones ES correctamente dirigidos, que retienen, por ejemplo, un número de copias de dos de mArm, de los clones en los que una delección inducida por Cas9 del locus objetivo de ratón 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	TGGGGCCGATCTTA GCC (SEQ ID NO: 14)	ACGAGGGGTTTCGG CCCATTC (SEQ ID NO: 15)	TTGACCGATTTCCTT GCGG (SEQ ID NO: 16)
HYG-U	LOA	CGAGTCTGTGAG AAGTTTCTG (SEQ ID NO: 17)	AGTTGACAGCGTIG TCCGACCTGA (SEQ ID NO: 18)	CAGCCCTCTTACA TCGAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGGGTACAC AAATCG (SEQ ID NO: 20)	CCGTCGGACCGAT GGCTGTGT (SEQ ID NO: 21)	GGGCGTCGGTTTCC ACTATC (SEQ ID NO: 22)
PGKp1 (Promotor Hyg)	LOA	CAAAATGGAAGTAGC ACGTCTACT (SEQ ID NO: 23)	CTCGTGCAGATGGA CAGACCCGC (SEQ ID NO: 24)	CCGCTGCCCAAG G (SEQ ID NO 25)
hlgK6	Número de copias de hArm1	GTCAAGCACGTGCTG GCACAC (SEQ ID NO: 26)	AACCCCTTGCTAT TGAATTGCTATGCT GTCAG (SEQ ID NO: 27)	TGTTGTAGACCCTC CQCCAC (SEQ ID NO: 28)
hlgK12 (inserto MAID 1710)	GOA	TGGCCTTCTCACA CCTGCAG (SEQ ID NO: 29)	CAGCCCATCCTGTC ACTTCGCTGGA (SEQ ID NO: 30)	TGGCCCAACAGTAC AGCTCAG (SEQ ID NO: 31)
hlgK13	Número de copias de hArm2	TCAGTCAATCACCT TTCCAGC (SEQ ID NO: 32)	TCCCAGGTAGCCT CATGAACCAATGTT (SEQ ID NO: 33)	CACATTACTGAGTC CCCACAGGG (SEQ ID NO: 34)
hlgK14	Número de copias de hArm2	CATTGTCAAAGAAG CACTGGAAATG (SEQ ID NO: 35)	ACCATTGCAAGTTTA CCCACGGTTAGGAT TTTT (SEQ ID NO: 36)	TCTTGCAATGGGAT CATCAGATG (SEQ ID NO: 37)

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
Neo	GOA	GGTGGAGAGGCTAT TCGGC (SEQ ID NO: 38)	TGGCACAAACAGAC AATCGGCTG (SEQ ID NO: 39)	GAACACGGCGGCAT CAG (SEQ ID NO: 40)
hlgK15	GOA	CAGGTGCAAAGGTG ACCACAG (SEQ ID NO: 41)	TGGGTCTGCCCAT CCATGCA (SEQ ID NO: 42)	GGCAGCCTGAGTGT CAGAGC (SEQ ID NO: 43)
hlgK25	GOA	GTTCAGGCCCCACA GACTCTC (SEQ ID NO: 44)	TCTCTCTGGAGCA ACCATGAAAGTTCCC T (SEQ ID NO: 45)	CCTGAAGCCATGAG GGCAG (SEQ ID NO: 46)
hUbc-D (Neo Promotor)	GOA	AGGGTAGGCTCTCC TGAATCG (SEQ ID NO: 47)	ACAGGCGCCGGACC TCTGGT (SEQ ID NO: 48)	CCAAAGAAACTGAC GCCTCAC (SEQ ID NO: 49)
TCRA Arm4	Número de copias de mArm	GGCCACATGAATT TGACCAG (SEQ ID NO: 50)	TGTACCCAAATCTC CAAAGAAAAGAGCT G (SEQ ID NO: 51)	GGCATCTGTCTC CCTTC (SEQ ID NO: 52)
1540m1 parental	Control de ratón de tipo silvestre	CAGTAAGGGAAGA GACTACAACAGCAT (SEQ ID NO: 53)	TGCACACTGCTCAC CACTGCAAGCTAT (SEQ ID NO: 54)	TGCTGGTGGCCCCA TCT (SEQ ID NO: 55)
1540m3 parental	Control de ratón de tipo silvestre	GAACTCAGCTATGA TAGTGTGGAATGTA (SEQ ID NO: 56)	CAGCCAGCAGCTG TGGGTCTC (SEQ ID NO: 57)	GCTCAGGGAACA CAGAACTTAGA (SEQ ID NO: 58)
hlgK5	Secuencia MAID 6548	CCCCGTCCTCTCC TTTTTC (SEQ ID NO: 59)	TCAATGCCAATTAAC CCATTTACCTTTTGC CCA (SEQ ID NO: 60)	TGCAAGTGTGCCA GCAAG (SEQ ID NO: 61)



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

5 Tabla 3: Secuencias de unión de locus resultantes de la transformación de células ES dobles

Unión	Seq ID No	Secuencia
Tcra/5' Frt de ratón	62	(GTCTTTTTGTTCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACACTCTTCA) <b>GAAGGGACTC</b> <b>CTGATTTCAAAGGGTACCGAAGTTCCTATT</b> <i>CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC</i>
3' Frt/IgK humana	63	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGAACTTCCTAGGGTTTCACCG</i> <b>GTGGCGGCCTAACAGAGAGGAAAGTCAAA</b> TTATAAAGAATATGAGATTCAGAATICTGA TTAACTGTGG
IgK humana/Tcra de ratón	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC <b>GCCACCGCGGTGGAGCTC</b> (AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

El tamizaje de modificación de alelos (MOA) de las colonias de células ES aisladas dio como resultado la identificación de 27 clones correctamente dirigidos entre 960 colonias tamizadas, para una eficacia de transformación del 2,81 %.

Una estrategia alternativa para generar los loci TCR A que comprende los segmentos de genes de región variable de inmunoglobulina adicionales implica la transformación en serie con vectores de transformación grandes sucesivos (ver, por ejemplo, la Figura 2). Como tal, las células ES heterocigotas para todos los segmentos del gen JK humano y cuatro segmentos del gen VK humano funcional (MAID 6548) se electroporaron con un gran vector de transformación que comprende, de 5' a 3': un brazo de homología 5' de ratón de 15,5 kb, un casete de selección Frt-Ub-Neo-Frt, un fragmento de 120 kb que comprende los segmentos de genes VK3-7 a VK3-15, y un brazo de homología humana 3' de 30 kb que comprende los segmentos de genes Vκ1-5 y Vκ1-6 (también presente en la secuencia MAID 6548). La incorporación exitosa se confirmó con los ensayos TAQMAN® descritos anteriormente, con el uso de cebadores y sondas que se enumeran en la Tabla 2 arriba e indicado en la Figura 2: Hyg, hlgK5, hlgK6, hlgK12, Neo, 1540m3 parental, 1540m1 parental. En particular, las sondas TCRA Arm4 y hlgK6 se usaron como sondas específicas de brazo para validar la transformación genómica correcta del LTVEC. Puede usarse además un conjunto adicional de cebadores y sonda, hlgK10, para confirmar la incorporación exitosa: Cebador directo - CGATTATGACTGGTTAGGTAGAAAGGTG (SEQ ID NO: 65); Sonda - GCCACTGGTTTCTCCAATGTTTTCAATCCAT (SEQ ID NO: 66); Cebador reverso - GGGAGTACTTGGAGATCCCTAAGC (SEQ ID NO: 67).

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

55 Tabla 4: Secuencia de unión del locus resultante de la transformación de células ES simple

Unión	Seq ID No	Secuencia
Tcra/5' Frt de ratón	68	(TTGAGCTTCATCAAAGTCACATGGGTAA CTCTATGGAGTAGTCAGAACACACTCTCA) <b>GAAGGGACTCCTGATTCAAAGGGTACC</b> <i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAA</i> <i>AGTATAGGAACTTC</i>
3' Frt/IgK humana	69	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGAACTTCCTAGGGTTTCACCG</i> <b>GTGGCGCGCCAGGACCCAGGCTCTGACACT</b> CAGGCTGCCAATACAATTGCCATGAAGACA GATGTTGATG
IgK humana/Tcra de ratón	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC <b>GCCACCGCGGTGGAGCTC</b> (AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

El tamizaje MOA de colonias aisladas dio como resultado la identificación de 5 clones correctamente dirigidos entre 440 colonias tamizadas (LTVEC solo), para una eficacia de transformación del 1,1 %. Los resultados para el tamizaje de colonias aisladas dirigidas con LTVEC + ZFN o LTVEC + CRISPR-Cas9 se muestran en la **Tabla 9**.

Después de completar la transformación única representada en la **Figura 2**, las células ES pueden dirigirse sucesivamente con grandes vectores de transformación que comprenden VK adicional para sumar a todo el repertorio de los segmentos de genes VK de la inmunoglobulina humana funcional.

En otras estrategias alternativas, la transformación doble o única de segmentos de genes VK de Ig humana sucesivos adicionales puede lograrse con el uso de los esquemas de transformación dobles (dos vectores de transformación grandes) o simples (un vector de transformación grande) que implican la destrucción mediada por nucleasas de dedos de zinc o CRISPR de un(os) casete(s) de selección (por ejemplo, higromicina).

Las células ES dirigidas descritas anteriormente se usan como células ES donadoras y se introducen en un embrión en la etapa pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células mediante el método VELOCIMOUSE® (ver, 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 donadoras se incuba hasta la etapa de blastocisto y después se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES donadoras. Los ratones F0 completamente derivados de la célula ES donante que llevan independientemente un gen quimérico IgK V humano- Tcra C de ratón se identifican mediante genotipado con el uso de una modificación del ensayo de alelos que detecta la presencia de secuencias genéticas únicas.

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

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

Se diseñaron varios ARN guía (ARNg) para reconocer varias secuencias objetivo dentro del gen de resistencia a la higromicina (secuencia de reconocimiento CRISPR). Las secuencias de reconocimiento CRISPR dentro del gen de higromicina fueron las siguientes: ARNg#1: ACGAGCGGGTTCGGCCCATTCGG (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 higromicina se representan en la **Figura 3**, que representa la destrucción mediada por CRISPR/Cas de la higromicina en el vector de transformación MAID 1545. Se tamizaron el ARNg#1, ARNg#6, ARNg#10 y ARNg#16 y se confirmaron que se dirigen específicamente al gen de higromicina (ver la **Figura 3**). Los resultados del tamizaje primario con el uso de los diversos ARNg específicos de la higromicina se proporcionan en la **Tabla 5**.

Tabla 5. Resultados del tamizaje primario con el uso de los ARNg específicos de la higromicina.

ARNg/ZFN	# total de candidatos primarios (2 placas/electroporación)	Candidatos positivamente reconfirmados / Total reconfirmado
gRNA1	5	2/2
gRNA6	6	1/1
gRNA10	19	5/5
gRNA16	91	8/8
ZFN 1/2	10	4/4

Células ES, por ejemplo, la célula ES generada en la Figura 1 (MAID 6548, heterocigoto para todos los segmentos JK humanos y cuatro segmentos de genes VK humanos funcionales) se electroporaron con dos vectores de transformación grandes (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 objetivo dentro del gen de resistencia a la higromicina.

Los dos vectores de transformación se insertaron mediante la recombinación homóloga en la secuencia de ADN, reemplazando la región que contiene y rodea el casete de selección de la higromicina. La incorporación exitosa de los dos vectores de transformación grandes se confirmó con el uso de ensayos TAQMAN®.

Las células ES dirigida descritas anteriormente se usarán como células ES donadoras y se introducen en un embrión en la etapa pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células mediante el método VELOCIMOUSE® (ver, 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 genéticamente modificadas se incubaba hasta la etapa de blastocisto y después se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES donadoras. Los ratones F0 completamente derivados de las células ES donadoras se identificarán mediante el genotipado con el uso de un ensayo de modificación de alelos que detecta la presencia de secuencias genéticas únicas.

Ejemplo 3: Transformación del locus TCR Alfa con tres LTVEC en combinación con las nucleasas con dedos de zinc

Se diseñó un sistema de transformación triple de manera que tres vectores de transformación grandes (LTVEC) modifiquen un locus genómico en una sola etapa de transformación. Como se muestra en la Figura 4, una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de transformación de la higromicina se dirigió por el sistema de transformación triple para generar las células ES que comprenden segmentos de genes variables IgK adicionales.

Este enfoque de triple transformación, se resume en la Figura 4, implica una triple transformación o coelectroporación de tres vectores de transformación grandes diferentes (LTVEC) (MAID 6647, MAID 6600 y MAID 1710) junto con las secuencias de nucleótidos que codifican una endonucleasa (por ejemplo, nucleasas con dedos de zinc o Cas9 y los ARNg) que crea una ruptura de doble cadena en o cerca del locus objetivo en las células ES.

En este enfoque, el primer vector de transformación grande (marcado como MAID 1710) comprendía un brazo de homología de 3' de 30 kb que incluye la secuencia de segmentos de genes humanos Vk1-5 y Vk1-6, una secuencia de 120 kb que comprendía los segmentos de genes Vk3-7 a Vk3-15 humanos, y una región 5' de 20 kb ("región superpuesta") que comprendía el segmento del gen Vk1-16 humano. El segundo vector de transformación grande (marcado como MAID 6600) comprendía una región de superposición 3' de 20 kb (región que comprende el segmento del gen Vk1-16 humano, igual que en el primer vector), una secuencia de 140 kb que comprende los segmentos de genes Vk1-17 a Vk2-24 y una región 5' de 60 kb ("región superpuesta") que comprendía Vk3-25 a Vk2-30 humano. El tercer vector de transformación grande (marcado como MAID 6647) comprendía una región de superposición de 3' de 60 kb (región que comprende Vk3-25 a Vk2-30 humano, igual que en el segundo vector), una secuencia de 90 kb que comprende Vk3-31 a Vk2-40 humano, un casete de transformación FRT-Ub-Neo-FRT y un brazo de homología 5' TCR A de ratón de 15,5 kb.

Las nucleasas con dedos de zinc (ZFN) se diseñaron para reconocer y escindir una secuencia objetivo dentro del gen de resistencia a la higromicina para promover la recombinación homóloga de tres LTVEC en el locus TCR A objetivo. Las células ES generadas en la Figura 4 (MAID 6548, heterocigoto para todos los segmentos Jk humanos y cuatro segmentos de genes VK humanos funcionales) se electroporaron con los tres vectores de transformación grandes (MAID6600 cortado, MAID1700 cortado y MAID6647 cortado) descritos anteriormente y dos plásmidos que expresan cada mitad de la ZFN (1/2), que se unen a las secuencias de reconocimiento en el gen de resistencia a la higromicina y catalizan una ruptura de doble cadena en el sitio objetivo (TGCGATCGCTGCGGCCGAtcttagCCAGACGAGCGGGTTCGG (SEQ ID NO: 2); con el sitio de escisión en letras minúsculas) (ver Tabla 1).

Los tres vectores de transformación sometidos a electroporación conjunta se insertaron mediante la recombinación homóloga en la secuencia de ADN, reemplazando la región que contiene y rodea el casete de selección de la higromicina. Las células ES resultantes contenían en el locus TCR A endógeno un dominio variable de la 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 transformación grandes se confirmó con el uso de los ensayos TAQMAN® descritos anteriormente (Lie y Petropoulos, 1998. Curr. Opin. Biotechnology 9:43-48), con el uso de sondas e cebadores indicados en la Figura 4 y enumerado en la Tabla 2 arriba y en Tabla 6 más abajo (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 vector de transformación grande (MAID 1710); hArm2 = superposición de 20 kb del primero (MAID 1710) y el segundo (MAID 6600) vectores de transformación grandes, hArm3 = 60 kb superpuesta del segundo (MAID 6600) y el tercer (MAID6647) vectores de transformación, mArm = brazo de homología 5' de 15,5 kb del tercer vector de transformación (MAID 6647), control WT de ratón - secuencias presentes en el locus TCR A de ratón). Los ensayos de PCR en tiempo real que reconocen las secuencias en los brazos de homología de los LTVEC, denominados ensayos específicos del brazo, se utilizaron para verificar la orientación correcta del LTVEC en el genoma del ratón. La determinación del número de copias de estos ensayos específicos del brazo proporcionó un aclaramiento adicional para ayudar a distinguir los clones ES correctamente dirigidos, que retienen, 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 los clones en los que una delección inducida por Cas9 del locus objetivo de ratón 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) para la sonda de ratón (mArm), y un número de copias de dos (o más) para la sonda humana (hArm1). Para detectar el ensamblaje correcto de los tres LTVEC por recombinación homóloga en el locus deseado, utilizamos ensayos TAQMAN® específicos del brazo. Los números de copia esperados, 1 para hArm2 y hArm3, indicaron que los tres LTVEC se han ensamblado dentro de un único constructo.

Tabla 6: Cebadores y sondas TAQMAN

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
HYG	LOA	TGGGGCCGATCTTAG CC (SEQ ID NO: 14)	ACGAGCGGGTTCGGCC CATTG (SEQ ID NO: 15)	TTGACCGATTCCTTGC GG (SEQ ID NO: 16)
HYG-U	LOA	CGACGTCTGTCGAGA AGTTTCTG (SEQ ID NO: 17)	AGTTCGACAGCGTGTG CGACCTGA (SEQ ID NO: 18)	CACGCCCTCCTACATC GAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGCGTACACA AATCG (SEQ ID NO: 20)	CCGTCTGGACCGATGG CTGTGT (SEQ ID NO: 21)	GGGCGTCGGTTTCCAC TATC (SEQ ID NO: 22)
hlgK6	Número de copias de hArm1	GTCAAGCACCTGCTGG CACAC (SEQ ID NO: 26)	AACCTTGTGCTATTGA ATTGCTATGCTGTGAG (SEQ ID NO: 27)	TGTTGTAGACCCTCCG CCAC (SEQ ID NO: 28)
hlgK12 (inserto MAID 1710)	GOA	TTGCCTTTCTCACACC TGCAG (SEQ ID NO: 29)	CAGCCCATCCTGTTCAC TTGCTGGA (SEQ ID NO:30)	TGGCCCAACAGTACA GCTCAG (SEQ ID NO: 31)
hlgK13	Número de copias de hArm2	TCAGTCAATCACCTTT CCCAGC (SEQ ID NO: 32)	TCCCCAGGTAGCCTCA TGAACCAAATGTT (SEQ ID NO: 33)	CACATTACTGAGTCCC CACAGGG (SEQ ID NO: 34)
hlgK14	Número de copias de hArm2	CATTGTCAAAGAAGC ACTGGAATG (SEQ ID NO: 35)	ACCAATTGCAGTTTACCC ACGGTTAGGATTTTT (SEQ ID NO: 36)	TCTTGCCAAATGGGATCA TCAGATG (SEQ ID NO: 37)
Neo	GOA			

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
		GGTGGAGAGGCTATT CGGC (SEQ ID NO: 38)	TGGGCACAACAGACAA TCGGCTG (SEQ ID NO: 39)	GAAACACGGCGGCATC AG (SEQ ID NO: 40)
hlgK15	GOA	CAGGTGCAAAAGGTGA CCACAG (SEQ ID NO: 41)	TGGGTCTGCCCCATCC ATGCA (SEQ ID NO: 42)	GGCAGCCTGAGTGTC AGAGC (SEQ ID NO: 43)
hlgK25	Número de copias de hArm3	GTTCAGGCCCCACACAG ACTCTC (SEQ ID NO: 44)	TCCTCTGGAGCAAC CATGAAGTTCCT (SEQ ID NO: 45)	CCTGAAGCCATGAGG GCAG (SEQ ID NO: 46)
TCRA-Arm4	Número de copias de mArm	GCGCCACATGAATTT GACCAG (SEQ ID NO: 50)	TGTACCCCAATCTCCAA AGAAAGAGCTG (SEQ ID NO: 51)	GGCATCCTGTCTCCC TTC (SEQ ID NO: 52)
1540m1 parental	Control de ratón de tipo silvestre	CAGTAAGGGAAGAG ACTACAACAGCAT (SEQ ID NO: 53)	TGCACACTGGTCAACCA CTGCAAGCTAT (SEQ ID NO: 54)	TGCTGGTGCCCCATC T (SEQ ID NO: 55)
1540m3 parental	Control de ratón de tipo silvestre	GAACCTCAGCTATGAT AGTGTGAAATGTA (SEQ ID NO: 56)	CAGCCCAGCAGCTGTG GGTTCTC (SEQ ID NO: 57)	GCTCAGGGAGAACAC AGAACTTAGA (SEQ ID NO: 58)
hlgK5	Secuencia MAID 6548 (ver Tabla 2)	CCCCGTCTCTCTCTT TTTC (SEQ ID NO: 59)	TCATGTCCATTAACCCA TTTACCTTTGGCCA (SEQ ID NO: 60)	TGCAAGTGTGCCAG CAAG (SEQ ID NO: 61)
hlgK22	Número de copias de hArm3			

Gen	Ensayo	Cebador directo	Sonda	Cebador inverso
		TGGCTCAAGAACAG TTTGCC (SEQ ID NO: 74)	CCCTGACTTTGCTGCTC AACTCACAGCC (SEQ ID NO: 75)	GGTCCAGTGGAAATCT GCCATG (SEQ ID NO: 76)
hlgK21	GOA	CATTTGGCTACATAT CAAAGCCG (SEQ ID NO: 77)	CCTGAGCCAGGGAACA GCCCACTGATA (SEQ ID NO: 78)	ACATGGCTGAGGCAG ACACC (SEQ ID NO: 79)
hlgK26	GOA	TGGGCCGTTATGCTA GTACCA (SEQ ID NO: 80)	TGGCTTACCCCTTTTG AAGGGCC (SEQ ID NO: 81)	CACAGCTGAAAGCAGG ATGAGC (SEQ ID NO: 82)
hlgK30	GOA	TCTCTGAGCAGCCAT CCCC (SEQ ID NO: 83)	TTCTCTTTGGTGTAGA GGCACCCAGC (SEQ ID NO: 84)	ACCAGGCATGGCAGA AAGG (SEQ ID NO: 85)

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

5 Tabla 7: Secuencias de unión de locus resultantes de la transformación de células ES triples

Unión	Seq ID No	Secuencia
Tcra/5' Frt de ratón	62	(GTCTTTTTTGTTCCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACACTCTTCA) <b>GAAGGGACTC</b> <b>CTGATTTCAAAGGGTACCGAAGTTCCTATT</b> <i>CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC</i>
3' Frt/IgK humana	86	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGAACTTCCTAGGGTTTCACCGGTG</i> <b>GCGCGCCTGAGTAGTGCTTTAGGTGTGAATCA</b> <i>CCAAAGATTTAGTGAAGTCCCTGTGCAAGGAG</i>
IgK humana/Tcra de ratón	64	GATAAATTATTTTGTGACAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC <b>GCCACCGCGGTGGAGCTC</b> (AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

El tamizaje de modificación de alelos (MOA) de las colonias de células ES aisladas resultaron en una eficacia de transformación de 0,4 % (ver Tabla 8).

35 Tabla 8. Resultados del tamizaje de modificación de alelos (MOA) para la transformación con 3 LTVEC

# de los LTVEC	Nucleasa	Delección	Inserción	Eficiencia
3 LTVECs	ZFN	hyg	370 kb humano	0,4 %
3 LTVECs	gRNA#16/Cas9	hyg	370 kb humano	0,4 %
3 LTVECs	ninguna	hyg	370 kb humano	0 %

45 Las células ES transformadas descritas anteriormente se usan como células ES donadoras y se introducen en un embrión en etapa pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células mediante el método VELOCIMOUSE® (ver, 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 donadoras se incuba hasta la etapa de blastocisto y después se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES donadoras. Los ratones F0 completamente derivados de la célula ES donante que llevan independientemente un gen quimérico IgK V humano- Tcra C de ratón se identifican mediante genotipado con el uso de un ensayo de modificación de alelos que detecta la presencia de secuencias genéticas únicas.

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

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

60 Se diseñaron varios ARN guía (ARNg) para reconocer varias secuencias objetivo 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: ACGAGCGGGTTCGGCCCATTCGG (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 higromicina se representan en la Figura 3. Se seleccionaron el ARNg#1, ARNg#6, ARNg#10 y ARNg#16 y confirmaron que apuntan específicamente al gen de la higromicina (ver Figura 3 y Tabla 5).



Las células ES MAID 6548 (heterocigotas para todos los segmentos de Jk humano y cuatro segmentos de genes VK humano funcionales) se electroporaron con tres vectores de transformación grandes como se describió en el Ejemplo 3, junto con los vectores que codifican Cas9 y ARNg#16, que reconocen y escinden un sitio objetivo dentro del gen de resistencia a la higromicina.

5 Los tres vectores de transformación grandes se insertaron mediante recombinación homóloga en la secuencia de ADN, reemplazando la región que contiene y que rodea el casete de selección de la higromicina. La incorporación exitosa de los tres vectores de transformación grandes se confirmó con el uso de los ensayos TAQMAN® descritos en el Ejemplo 3.

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

15 El tamizaje de modificación de alelos (MOA) de las colonias de células ES aisladas dio como resultado una eficacia de transformación de 0,4 % (ver Tabla 8).

20 Las células ES dirigida descritas anteriormente se usarán como células ES donadoras y se introducen en un embrión en la etapa pre-mórula, por ejemplo, un embrión de ratón en etapa de 8 células mediante el método VELOCIMOUSE® (ver, 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 genéticamente modificadas se incuba hasta la etapa de blastocisto y después se implanta en una madre sustituta para producir un ratón F0 completamente derivado de las células ES donadoras. Los ratones F0 completamente derivados de las células ES donadoras se identificarán mediante el genotipado con el uso de una modificación del ensayo de alelos que detecta la presencia de secuencias genéticas únicas.

25 Ejemplo 5: Mejoramiento de la transformación de LTVEC a través de secuencias de superposición entre los dos LTVEC

30 El sistema de doble transformación descrito en el Ejemplo 1 se empleó para modificar un locus genómico en una sola etapa de transformación con el uso de dos vectores de transformación grandes (LTVEC). Como se muestra en la Figura 1, una célula que tiene una modificación heterocigótica del locus alfa TCR en el cromosoma 14 de ratón que comprende un casete de selección de la higromicina se dirigió por el sistema de doble transformación para generar las células ES que comprenden los segmentos de genes variables de IgK adicionales. Los dos LTVEC diferentes se coelectroporaron juntos en las células madre embrionarias de ratón (ES). Opcionalmente, un ácido nucleico que codifica una endonucleasa (ya sea una nucleasa de dedos de zinc (ZFN) o CRISPR-Cas9) se coelectropora para crear una ruptura de doble cadena en o cerca del locus objetivo.

35 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 segmentos de genes VK1-5 y VK1-6 humanos, una secuencia de 120 kb que comprendía los segmentos de genes VK3-7 a VK3-15 humanos y una región 5' de 20 kb ("región superpuesta") que comprendía un segmento genético VK1-16 humano. El segundo LTVEC (marcado como MAID 6600) comprendía una región de superposición 3' de 20 kb (región que comprende el segmento del gen VK1-16 humano, igual que en el primer vector), una secuencia de 140 kb que comprende los segmentos de genes Vκ1-17 a Vκ2-30 humanos, un casete de selección FRT-Ub-Neo-FRT y un brazo de homología 3' TCR A de ratón de 15,5 kb.

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

50 Como comparación, el único sistema LTVEC descrito en el Ejemplo 1 se empleó además para modificar el mismo locus genómico con el uso de un único LTVEC, ya sea solo o en combinación con una ZFN o CRISPR-Cas9 (ver Figura 2). La incorporación exitosa se confirmó con los ensayos TAQMAN descritos anteriormente, con el uso de cebadores y sondas que se enumeran en la Tabla 2 arriba e indican en la Figura 2

55 La Tabla 9 compara las eficiencias de transformación en los experimentos de transformación con el uso del LTVEC único (solo, con ZFN o con Cas9), con el uso de los dos LTVEC simultáneamente (solo, con ZFN o con Cas9), o con el uso de los dos LTVEC más un tercer LTVEC simultáneamente (solo, con ZFN o con Cas9). Las eficiencias de transformación presentadas en la Tabla 9 son el porcentaje de clones ESC tamizados que se determinó que estaban correctamente dirigidos a través del tamizaje inicial, el tamizaje de confirmación y el tamizaje de reconfirmación con el uso de los cebadores y sondas TAQMAN Tabla 2. La transformación con un solo LTVEC solo dio como resultado un 1,1 % de clones correctamente dirigidos. La escisión con una ZFN aumentó la eficiencia de transformación del LTVEC único a 4,4 %, y la escisión con CRISPR-Cas9 aumentó la eficiencia de transformación del LTVEC único a 5,5 %. Sorprendentemente, la transformación con los 2 LTVEC que tienen 20 kb en la secuencia de superposición dio como resultado una eficiencia de transformación del 1,4 % incluso cuando no se usó la nucleasa. La eficacia de la transformación aumentó a 2,81 % cuando se usó una ZFN y 1,6 % cuando se usó Cas9.

Tabla 9: Eficiencia de transformación de los LTVEC coelectroporados optimizados

	MAID 6598: 2 <sup>da</sup> Inserción		MAID 6600: 2 <sup>da</sup> y 3 <sup>ra</sup> Inserciones		MAID 6647: 2 <sup>da</sup> , 3 <sup>ra</sup> , y 4 <sup>ta</sup> Inserciones			
	LTVEC solo	LTVEC + ZFN (4,4 / 384 (1,1 %))	LTVEC + Cas9 (21 / 384 (5,5 %))	2 LTVEC solos (13 / 960 (1,4 %))	2 LTVEC + ZFN (27 / 960 (2,8 %))	2 LTVEC + Cas9 (15 / 960 (1,6 %))	3 LTVEC + ZFN (4 / 960 (0,4 %))	3 LTVEC + Cas9 (4 / 960 (0,4 %))
Eficiencia de transformación	5 / 440 (1,1 %)	17 / 384 (4,4 %)	21 / 384 (5,5 %)	13 / 960 (1,4 %)	27 / 960 (2,8 %)	15 / 960 (1,6 %)	4 / 960 (0,4 %)	4 / 960 (0,4 %)

Listado de secuencias

- 5 <110> Voronina, Vera  
 Macdonald, Lynn  
 Zambrowicz, Brian  
 Murphy, Andrew J.
- 10 <120> Métodos y composiciones para la modificación genética dirigida a través de la transformación múltiple en una sola etapa  
 <130> 57766-472224
- 15 <150> US 62/094,104  
 <151> 2014-12-19
- 20 <150> US 62/167,408  
 <151> 2015-05-28
- <150> US 62/205,524  
 <151> 2015-08-14
- <160> 86
- <170> FastSEQ para Windows Versión 4.0
- 25 <210> 1  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial
- 30 <220>  
 <223> una secuencia genómica objetivo que se une a un ARN guía (ARNg)
- 35 <220>  
 <221> misc\_característica  
 <222> (2)...(21)  
 <223> n = A,T,C or G
- 40 <400> 1  
 gnnnnnnnnn nnnnnnnnn ngg 23
- 45 <210> 2  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial
- 50 <220>  
 <223> sitio de reconocimiento ZFN1/2
- <400> 2  
 tgcgatcgct gcggccgatc ttagccagac gagcgggttc gg 42
- 55 <210> 3  
 <211> 36  
 <212> ADN  
 <213> Secuencia artificial
- 60 <220>  
 <223> sitio de reconocimiento de ZFN3/4
- <400> 3  
 cgctgcggcc gatcttagcc agacgagcgg gttcgg 36
- 65 <210> 4  
 <211> 39  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> sitio de reconocimiento de ZFN5/6  
  
 <400> 4  
 5 agcgtgtccg acctgatgca gctctcggag gcggaagaa 39  
  
 <210> 5  
 <211> 18  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de unión a ZFN1  
  
 <400> 5  
 15 tgcgatcgct gcggccga 18  
  
 <210> 6  
 <211> 18  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de unión a ZFN2  
  
 <400> 6  
 25 ccgaaccgc tcgtctgg 18  
  
 <210> 7  
 <211> 15  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de unión a ZFN3  
  
 <400> 7  
 cgctgcggcc gatct 15  
  
 <210> 8  
 <211> 15  
 40 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de unión a ZFN4  
  
 <400> 8  
 50 ccgaaccgc tcgtc 15  
  
 <210> 9  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Secuencia de unión a ZFN5  
  
 <400> 9  
 60 agcgtgtccg acctgatg 18  
  
 <210> 10  
 <211> 15  
 <212> ADN  
 65 <213> Secuencia artificial

<220>  
 <223> Secuencia de unión a ZFN6  
  
 5 <400> 10  
 ttctcgccc tccga 15  
  
 <210> 11  
 <211> 6  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> sitio de escisión para ZFN1/2  
  
 15 <400> 11  
 tcttag 6  
  
 <210> 12  
 <211> 6  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sitio de escisión para ZFN3/4  
 25 <400> 12  
 tagcca 6  
  
 <210> 13  
 <211> 6  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sitio de escisión para ZFN5/6  
 35 <400> 13  
 cagctc 6  
  
 40 <210> 14  
 <211> 17  
 <212> ADN  
 <213> Secuencia artificial  
  
 45 <220>  
 <223> cebador directo HYG  
  
 <400> 14  
 50 tgcggccgat ctagcc 17  
  
 <210> 15  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 55 <220>  
 <223> sonda HYG  
  
 <400> 15  
 60 acgagcgggt tcggccatt c 21  
  
 <210> 16  
 <211> 18  
 <212> ADN  
 65 <213> Secuencia artificial

<220>  
 <223> cebador reverso HYG  
  
 <400> 16  
 5 ttgaccgatt ccttgagg 18  
  
 <210> 17  
 <211> 23  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo HYG-U  
  
 <400> 17  
 15 cgacgtctgt cgagaagttt ctg 23  
  
 <210> 18  
 <211> 24  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sonda HYG-U  
 25  
 <400> 18  
 agttcgacag cgtgtccgac ctga 24  
  
 <210> 19  
 <211> 19  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador reverso HYG-U  
  
 <400> 19  
 35 cacgccctcc tacatcgaa 19  
  
 <210> 20  
 <211> 20  
 40 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo HYG-D  
  
 <400> 20  
 45 tgcgggctg acacaaatcg 20  
 50  
 <210> 21  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <223> sonda HYG-D  
  
 <400> 21  
 60 ccgtctggac cgatggctgt gt 22  
  
 <210> 22  
 <211> 20  
 <212> ADN  
 65 <213> Secuencia artificial

<220>  
 <223> cebador reverso HYG-D  
  
 <400> 22  
 5 gggcgtcggg ttccactatc 20  
  
 <210> 23  
 <211> 24  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo PGKp1 (Promotor Hyg)  
  
 <400> 23  
 15 caaatggaag tagcacgtct cact 24  
  
 <210> 24  
 <211> 23  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sonda PGKp1 (Promotor Hyg)  
 25  
 <400> 24  
 ctctgtcaga tggacagcac cgc 23  
  
 <210> 25  
 <211> 15  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 35 <223> cebador reverso PGKp1 (Promotor Hyg)  
  
 <400> 25  
 ccgctgcccc aaagg 15  
  
 <210> 26  
 <211> 20  
 40 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 45 <223> cebador directo hlgK6  
  
 <400> 26  
 50 gtcaagcact gctggcacac 20  
  
 <210> 27  
 <211> 33  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <223> sonda hlgK6  
  
 <400> 27  
 60 aaccctgtg ctattgaatt gctatgctgt cag 33  
  
 <210> 28  
 <211> 20  
 <212> ADN  
 65 <213> Secuencia artificial

<220>  
 <223> cebador reverso hlgK6  
  
 <400> 28  
 5 tgtgtagac cctccgccac 20  
  
 <210> 29  
 <211> 21  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo hlgK12 (inserto MAID 1710)  
  
 <400> 29  
 15 ttgccttct cacacctgca g 21  
  
 <210> 30  
 <211> 25  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sonda hlgK12 (inserto MAID 1710)  
 25  
 <400> 30  
 cagcccatcc tgcactcg ctgga 25  
  
 <210> 31  
 <211> 21  
 <212> ADN  
 30 <213> Secuencia artificial  
  
 <220>  
 <223> cebador reverso hlgK12 (inserto MAID 1710)  
  
 <400> 31  
 tggccaaca gtacagctca g 21  
  
 <210> 32  
 <211> 22  
 <212> ADN  
 40 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo hlgK13  
  
 <400> 32  
 50 tcagtcaatc accttcca gc 22  
  
 <210> 33  
 <211> 28  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <223> sonda hlgK13  
  
 <400> 33  
 60 tccccaggta gcctcatgaa ccaatgtt 28  
  
 <210> 34  
 <211> 23  
 <212> ADN  
 65 <213> Secuencia artificial



<220>  
 <223> cebador reverso hlgK13  
  
 5 <400> 34  
 cacattactg agtccccaca ggg 23  
  
 <210> 35  
 <211> 25  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo hlgK14  
  
 15 <400> 35  
 cattgtcaaa gaagcactgg aaatg 25  
  
 <210> 36  
 <211> 32  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sonda hlgK14  
  
 25 <400> 36  
 accattgcag ttaccacag gtaggattt tt 32  
  
 <210> 37  
 <211> 23  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador reverso hlgK14  
  
 35 <400> 37  
 tcttgaatg ggatcatcag atg 23  
  
 40 <210> 38  
 <211> 19  
 <212> ADN  
 <213> Secuencia artificial  
  
 45 <220>  
 <223> cebador directo Neo  
  
 <400> 38  
 50 ggtggagagg ctattcggc 19  
  
 <210> 39  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
  
 55 <220>  
 <223> Sonda Neo  
  
 <400> 39  
 60 tgggcacaac agacaatcgg ctg 23  
  
 <210> 40  
 <211> 17  
 <212> ADN  
 65 <213> Secuencia artificial

ES 2 760 508 T3

<220>  
 <223> cebador reverso Neo  
  
 5 <400> 40  
 gaacacggcg gcatcag 17  
  
 <210> 41  
 <211> 21  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo hlgK15  
  
 15 <400> 41  
 caggtgcaaa ggtgaccaca g 21  
  
 <210> 42  
 <211> 21  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> hlgK15 sonda  
 25 <400> 42  
 tgggtcctgc ccatccatgc a 21  
  
 <210> 43  
 <211> 20  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador reverso hlgK15  
  
 <400> 43  
 ggcagcctga gtgtcagagc 20  
  
 40 <210> 44  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
  
 45 <220>  
 <223> cebador directo hlgK25  
  
 <400> 44  
 50 gttcaggccc cacagactct c 21  
  
 <210> 45  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
  
 55 <220>  
 <223> sonda hlgK25  
  
 <400> 45  
 60 tcctctctgg agcaacctg aagttccct 29  
  
 <210> 46  
 <211> 19  
 <212> ADN  
 65 <213> Secuencia artificial

# ES 2 760 508 T3

<220>  
<223> cebador reverso hlgK25

5 <400> 46  
cctgaagcca tgagggcag 19

<210> 47  
<211> 21  
<212> ADN  
10 <213> Secuencia artificial

<220>  
<223> cebador directo hUbC-D (Neo Promotor)

15 <400> 47  
agggtaggct ctctgaatc g 21

<210> 48  
<211> 20  
20 <212> ADN  
<213> Secuencia artificial

<220>  
<223> sonda hUbC-D (Neo Promotor)

25 <400> 48  
acaggcgccg gacctctggt 20

<210> 49  
30 <211> 21  
<212> ADN  
<213> Secuencia artificial

<220>  
35 <223> cebador reverso hUbC-D (Neo Promotor)

<400> 49  
ccaaagaaac tgacgcctca c 21

40 <210> 50  
<211> 21  
<212> ADN  
<213> Secuencia artificial

45 <220>  
<223> cebador directo TCRA Arm4

<400> 50  
50 ggccacatg aattgacca g 21

<210> 51  
<211> 28  
<212> ADN  
<213> Secuencia artificial

55 <220>  
<223> sonda TCRA brazo4

<400> 51  
60 tgtaccaat ctccaaaga aagagctg 28

<210> 52  
<211> 19  
<212> ADN  
65 <213> Secuencia artificial

ES 2 760 508 T3

<220>  
 <223> cebador reverso TCRA brazo4  
  
 <400> 52  
 5 ggcacacctgt cctcccttc 19  
  
 <210> 53  
 <211> 27  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo parental 1540m1  
  
 <400> 53  
 15 cagtaaggga agagactaca acagcat 27  
  
 <210> 54  
 <211> 27  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> sonda parental 1540m1  
 25  
 <400> 54  
 tgcacactgc tcaccactgc aagctat 27  
  
 <210> 55  
 <211> 17  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador reverso parental 1540m1  
  
 <400> 55  
 tgctggtggc cccatct 17  
  
 <210> 56  
 <211> 28  
 40 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> cebador directo parental 1540m3  
  
 <400> 56  
 50 gaactcagct atgatagtgt cgaatgta 28  
  
 <210> 57  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <223> sonda parental 1540m3  
  
 <400> 57  
 60 cagcccagca gctgtgggtt ctc 23  
  
 <210> 58  
 <211> 25  
 <212> ADN  
 65 <213> Secuencia artificial

ES 2 760 508 T3

<220>  
 <223> cebador reverso parental 1540m3

5 <400> 58  
 gctcaggag aacacagaac ttaga 25

<210> 59  
 <211> 20  
 <212> ADN  
 10 <213> Secuencia artificial

<220>  
 <223> cebador directo hlgK5

15 <400> 59  
 ccccgctctc ctccttttc 20

<210> 60  
 <211> 32  
 20 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> sonda hlgK5

25 <400> 60  
 tcatgtccat taaccattt acctttgccc ca 32

<210> 61  
 <211> 19  
 30 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> cebador reverso hlgK5

35 <400> 61  
 tgcaagtgt gccagcaag 19

40 <210> 62  
 <211> 156  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> secuencia de empalme Tcra de ratón/5' Frt

<400> 62  
 gtcttttttg ttcttcacag ttgagcttca tcaaagtcac atgggttaaa ctctatggag 60  
 tagtcagaac acactcttca gaagggactc ctgatttcaa aggtaccga agttcctatt 120  
 ccgaagtcc tattctctag aaagtatagg aacttc 156

50 <210> 63  
 <211> 132  
 <212> ADN  
 <213> Secuencia artificial

55 <220>  
 <223> secuencia de empalme de 3' Frt/ IgK humana

<400> 63  
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctt agggtttcac 60  
 cgtggcgcg cctaacagag aggaaagtca aattataaag aatatgagat tcagaattct 120  
 60 gattaactgt gg 132

<210> 64

ES 2 760 508 T3

<211> 124  
 <212> ADN  
 <213> Secuencia artificial

5 <220>  
 <223> secuencia de empalme IgK Humana/Tcra de ratón

<400> 64  
 gataaattat tttgtcagac aacaataaaa atcaatagca cggcctaaga gcgggcgcca 60  
 ccgcggtgga gctcaggttt ccggtactta acaacagagc acagatttag tggtagagga 120  
 ctct 124

10 <210> 65  
 <211> 28  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <223> cebador directo hlgK10

<400> 65  
 cgattatgac tggtaggta gaaaggtg 28

<210> 66  
 <211> 32  
 <212> ADN  
 <213> Secuencia artificial

25 <220>  
 <223> sonda hlgK10

<400> 66  
 gccactggtt tctccaaatg tttcaatcc at 32

<210> 67  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <223> cebador reverso hlgK10

<400> 67  
 gggagtactt ggagatccct aagc 24

<210> 68  
 <211> 136  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> secuencia de empalme Tcra de ratón/ 5' Frt

<400> 68  
 ttgagcttca tcaaagtcac atgggttaaa ctctatggag tagtcagaac acactcttca 60  
 gaagggactc ctgatttcaa aggtaccga agttcctatt ccgaagtcc tattctctag 120  
 aaagtatagg aacttc 136

55 <210> 69  
 <211> 132  
 <212> ADN  
 <213> Secuencia artificial

60 <220>  
 <223> secuencia de empalme de 3' Frt/ IgK humana

# ES 2 760 508 T3

```

<400> 69
gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctc agggtttcac 60
cggtaggcgcg ccaggaccca ggctctgaca ctcaggctgc caatacaatt gccatgaaga 120
cagatgttga tg 132

5 <210> 70
<211> 23
<212> ADN
<213> Secuencia artificial

10 <220>
<223> ARNg#1 CRISPR

<400> 70
acgagcgggt tcggccatt cgg 23

15 <210> 71
<211> 23
<212> ADN
<213> Secuencia artificial

20 <220>
<223> secuencia de reconocimiento ARNg#6 CRISPR

<400> 71
cttagccaga cgagcgggtt cgg 23

25 <210> 72
<211> 23
<212> ADN
<213> Secuencia artificial

30 <220>
<223> secuencia de reconocimiento ARNg#10 CRISPR

<400> 72
gccgatcta gccagacgag cgg 23

35 <210> 73
<211> 23
<212> ADN
<213> Secuencia artificial

40 <220>
<223> secuencia de reconocimiento ARNg#16 CRISPR

45 <400> 73
cgacctgatg cagctctcg agg 23

50 <210> 74
<211> 21
<212> ADN
<213> Secuencia artificial

55 <220>
<223> cebador directo hlgK22

<400> 74
tggctccaag aacagtttc c 21

60 <210> 75
<211> 28
<212> ADN
<213> Secuencia artificial

<220>

```

<223> sonda hlgK22  
 <400> 75  
 5 ccctgacttt gctgctcaac tcacagcc 28  
 <210> 76  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> cebador reverso hlgK22  
 <400> 76  
 15 ggtccagtgg aatctgcat g 21  
 <210> 77  
 <211> 23  
 <212> ADN  
 20 <213> Secuencia artificial  
 <220>  
 <223> cebador directo hlgK21  
 25 <400> 77  
 cattggcta catatcaaag ccg 23  
 <210> 78  
 <211> 27  
 <212> ADN  
 30 <213> Secuencia artificial  
 <220>  
 <223> sonda hlgK21  
 35 <400> 78  
 cctgagccag ggaacagccc actgata 27  
 <210> 79  
 <211> 20  
 <212> ADN  
 40 <213> Secuencia artificial  
 <220>  
 45 <223> cebador reverso hlgK21  
 <400> 79  
 acatggctga ggcagacacc 20  
 50 <210> 80  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 55 <220>  
 <223> cebador directo hlgK26  
 <400> 80  
 60 tgggccgtta tgctagtacc a 21  
 <210> 81  
 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
 65 <220>



ES 2 760 508 T3

<223> sonda hlgK26  
 <400> 81  
 5 tggcttacc cctttgaag ggccc 25  
 <210> 82  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> cebador reverso hlgK26  
 <400> 82  
 15 cacagctgaa gcaggatgag c 21  
 <210> 83  
 <211> 19  
 <212> ADN  
 20 <213> Secuencia artificial  
 <220>  
 <223> cebador directo hlgK30  
 25 <400> 83  
 tctctgagca gccatcccc 19  
 <210> 84  
 <211> 27  
 <212> ADN  
 30 <213> Secuencia artificial  
 <220>  
 <223> sonda hlgK30  
 35 <400> 84  
 ttctccttg gtgtagagg caccagc 27  
 <210> 85  
 <211> 19  
 <212> ADN  
 40 <213> Secuencia artificial  
 <220>  
 45 <223> cebador reverso hlgK30  
 <400> 85  
 accaggcatg gcagaaagg 19  
 50 <210> 86  
 <211> 130  
 <212> ADN  
 <213> Secuencia artificial  
 55 <220>  
 <223> unión 3' Frt/ IgK humana  
 <400> 86  
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttct agggtttcac 60  
 cgtggcgcg cctgagtagt gcttaggtg tgtaatcacc aaagatttag tgaagtcct 120  
 60 gtgcaaggag 130

**REIVINDICACIONES**

1. Un método para modificar un locus genómico objetivo en una célula de mamífero, que comprende:
  - (a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro del locus genómico objetivo;
  - (b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de 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 es de 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 donde el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de al menos 100 kb, en donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición 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 objetivo, en donde la primera secuencia de superposición es de al menos 1 kb, y en donde el locus genómico objetivo se modifica mediante la integración del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico entre los segmentos genómicos correspondientes; y
  - (c) seleccionar una célula transformada que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico integrado en el locus genómico objetivo.
  
2. El método de la reivindicación 1, en donde el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico comprenden el ADN genómico, opcionalmente en donde 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 reportero, uno o más casetes de expresión o un ácido nucleico flanqueado por las secuencias objetivo de recombinación específicas de sitio, opcionalmente en donde el ADN genómico es homólogo u ortólogo a una secuencia que se dirige a la delección en el locus genómico objetivo, y opcionalmente, en donde 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.
  
3. El método de la reivindicación 1 o 2, en donde:
  - (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; y/o
  - (II) el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos son ácidos nucleicos humanos.
  
4. El método de cualquiera de las reivindicaciones 1-3, en donde:
  - (I) el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es de aproximadamente 100 kb a 500 kb;
  - (II) el tamaño combinado del primer inserto de ácido nucleico y el segundo inserto de ácido nucleico es al menos 200 kb; o
  - (III) la célula transformada comprende el ADN genómico que comprende el primer inserto de ácido nucleico y el segundo inserto de ácido nucleico juntos, que tienen un tamaño combinado en el intervalo de 100 kb a 500 kb.
  
5. El método de cualquiera de las reivindicaciones 1-4, en donde:
  - (I) la primera secuencia de superposición del primer LTVEC es idéntica a la primera secuencia de superposición del segundo LTVEC; y/o
  - (II) el tamaño de la primera secuencia de superposición es de 1 kb a 70 kb; y/o
  - (III) el tamaño de la primera secuencia de superposición es de al menos 10 kb, opcionalmente en donde el tamaño de la primera secuencia de superposición es de al menos 20 kb.
  
6. El método de una cualquiera de las reivindicaciones 1-5, en donde la integración del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, o ambos en el locus genómico objetivo da como resultado uno o más de:
  - (I) una adición de una secuencia exógena en el locus genómico objetivo;
  - (II) una delección de una secuencia endógena en el locus genómico objetivo, opcionalmente en donde la delección es de 5 kb a 800 kb; y
  - (III) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de estos.
  
7. El método de cualquiera de las reivindicaciones 1-6, en donde:
  - (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

(II) el primer LTVEC es al menos 50 kb y el segundo LTVEC es al menos 50 kb, opcionalmente en donde el primer LTVEC tiene al menos 100 kb y el segundo LTVEC tiene al menos 100 kb.

- 5 8. Un método para modificar un locus genómico objetivo en una célula de mamífero, que comprende:
- 10 (a) introducir en la célula un agente tipo nucleasa que produce una ruptura de simple o doble cadena dentro del locus genómico objetivo;
- 15 (b) introducir en la célula un primer vector de transformación grande (LTVEC) que es de 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 es de 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 es de 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 donde el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico es de al menos 100 kb,
- 20 en donde el primer brazo de homología 3' del primer LTVEC tiene una primera secuencia de superposición homóloga al segundo brazo de homología 5' del segundo LTVEC, el segundo brazo de homología 3' del segundo LTVEC tiene una segunda secuencia de superposición 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 objetivo, en donde la primera secuencia de superposición es de al menos 1 kb y la segunda secuencia de superposición es de al menos 1 kb, y en donde el locus genómico objetivo 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
- 25 (c) seleccionar una célula transformada que comprende el primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico integrado en el locus genómico objetivo.
- 30 9. El método de la reivindicación 8, en donde el primer inserto de ácido nucleico y/o el segundo inserto de ácido nucleico y/o el tercer inserto de ácido nucleico comprenden el ADN genómico, opcionalmente en donde 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 reportero, uno o más casetes de expresión, o un ácido nucleico flanqueado por las secuencias objetivo de recombinación específicas de sitio, opcionalmente en donde el ADN genómico es homólogo u ortólogo a una secuencia que se dirige a la delección en el locus genómico objetivo, y
- 35 opcionalmente, en donde 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 humano con una secuencia de ácido nucleico humano homólogo u ortólogo.
- 40 10. El método de la reivindicación 8 o 9, en donde:
- (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; y/o
- 45 (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.
- 50 11. El método de cualquiera de las reivindicaciones 8-10, en donde:
- (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 aproximadamente 100 kb a 700 kb;
- (II) el tamaño combinado del primer inserto de ácido nucleico, el segundo inserto de ácido nucleico, y el tercer inserto de ácido nucleico es de al menos 200 kb; o
- (III) la célula transformada comprende el 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 en el intervalo de 100 kb a 700 kb.
- 55 12. El método de cualquiera de las reivindicaciones 8-11, en donde:
- (I) la primera secuencia de superposición del primer LTVEC es idéntica a la primera secuencia de superposición del segundo LTVEC, y/o la segunda secuencia de superposición del segundo LTVEC es idéntica a la segunda secuencia de superposición del tercer LTVEC; y/o
- 60 (II) el tamaño de la primera secuencia de superposición es de 1 kb a 70 kb, y/o el tamaño de la segunda secuencia de superposición es de 1 kb a 70 kb; y/o
- (III) el tamaño de la primera secuencia de superposición es de al menos 10 kb, y/o el tamaño de la segunda secuencia de superposición es de al menos 10 kb, opcionalmente en donde el tamaño de la primera secuencia de superposición es de al menos 20 kb, y/o el tamaño de la segunda secuencia de superposición es de al menos 20 kb.

13. El método de cualquiera de las reivindicaciones 8-12, en donde 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 objetivo da como resultado uno o más de:
- 5 (I) una adición de una secuencia exógena en el locus genómico objetivo;  
(II) una delección de una secuencia endógena en el locus genómico objetivo, opcionalmente en donde la delección es de 5 kb a 800 kb; o  
(III) una activación, una desactivación, una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de estos.
- 10 14. El método de cualquiera de las reivindicaciones 8-13, en donde:  
(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  
15 (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 donde 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.
- 20 15. El método de cualquiera de las reivindicaciones 1-14, en donde el agente tipo nucleasa es una nucleasa de dedos de zinc (ZFN), una nucleasa efectora de tipo activador de la transcripción (TALEN), o una meganucleasa.
16. El método de cualquiera de las reivindicaciones 1-15, en donde el agente tipo nucleasa es una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas regularmente Intercaladas (CRISPR) y un ARN guía (ARNg), opcionalmente en donde la proteína Cas es Cas9.
- 25 17. El método de cualquiera de las reivindicaciones 1-16, en donde la célula es:  
(I) una célula humana;  
(II) una célula no humana;  
(III) una célula de roedor, opcionalmente en donde el roedor es un ratón o una rata;  
30 (IV) una célula pluripotente, opcionalmente en donde la célula pluripotente es una célula madre embrionaria (ES) o una célula madre pluripotente inducida (iPS);  
(V) una célula madre hematopoyética;  
(VI) una célula madre neuronal; o  
(VII) una célula de fibroblastos.
- 35 18. El método de la reivindicación 17, en donde la célula es una célula ES de ratón o una célula ES de rata, que comprende opcionalmente además introducir la célula ES de ratón o la célula ES de rata en un embrión huésped no humano.

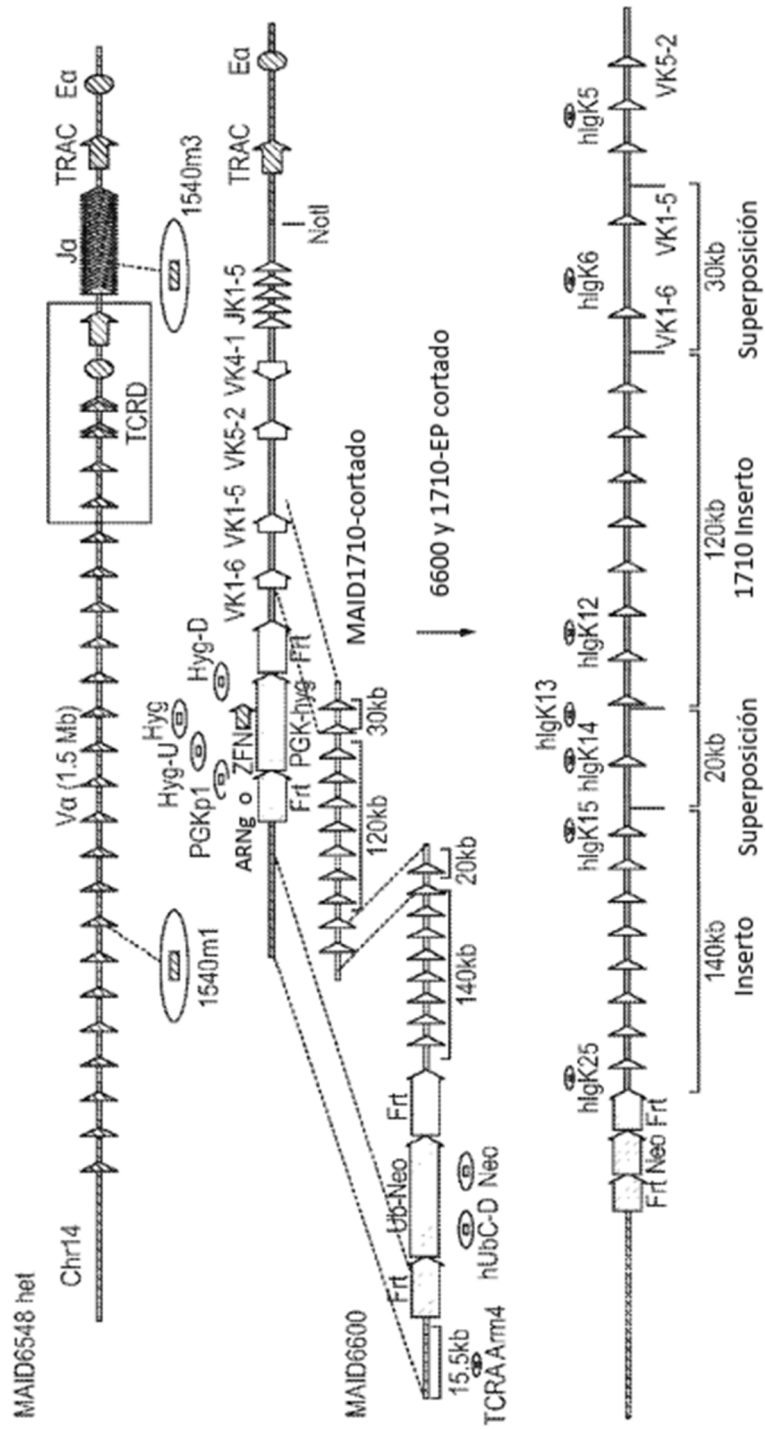


Figura 1

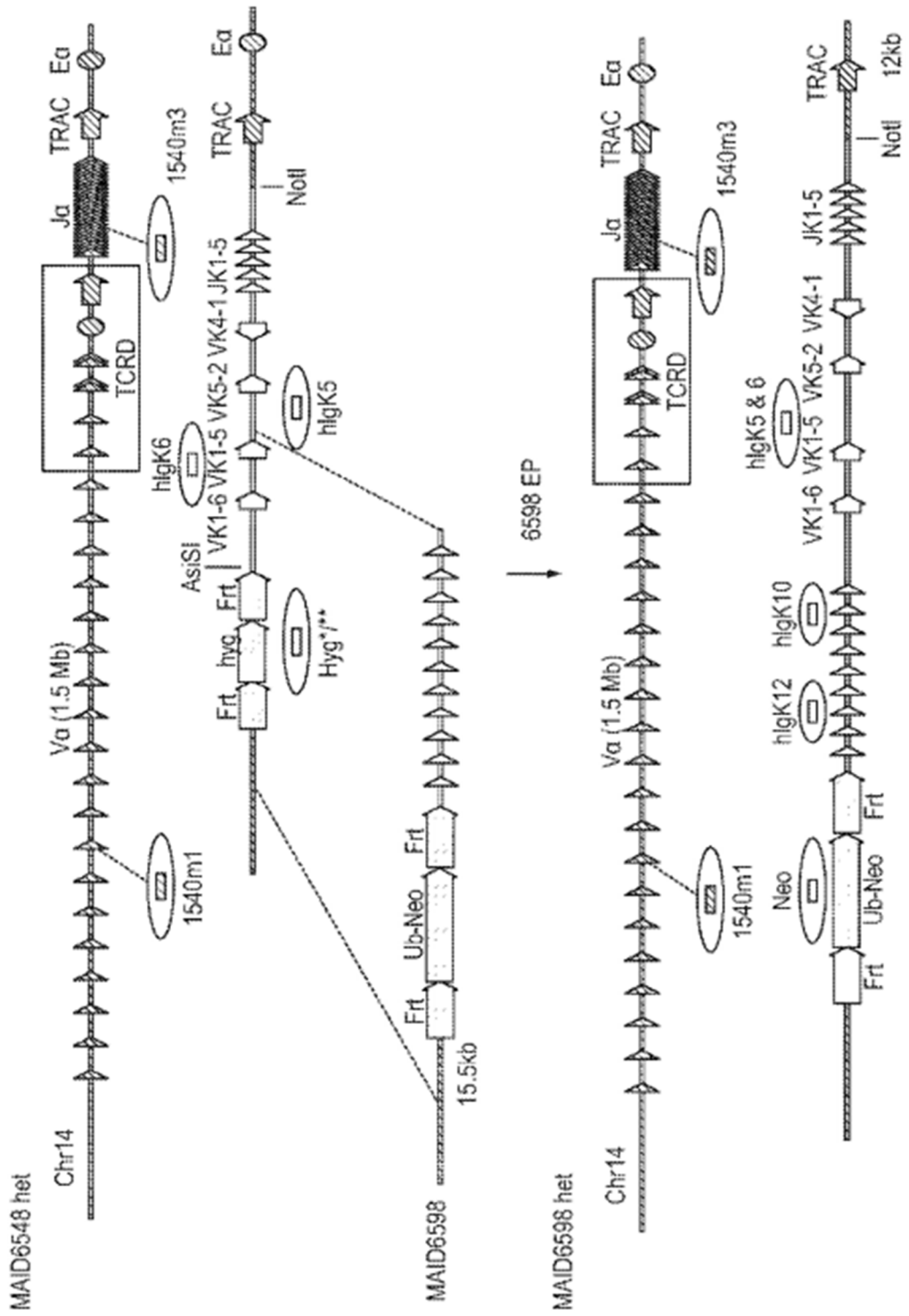
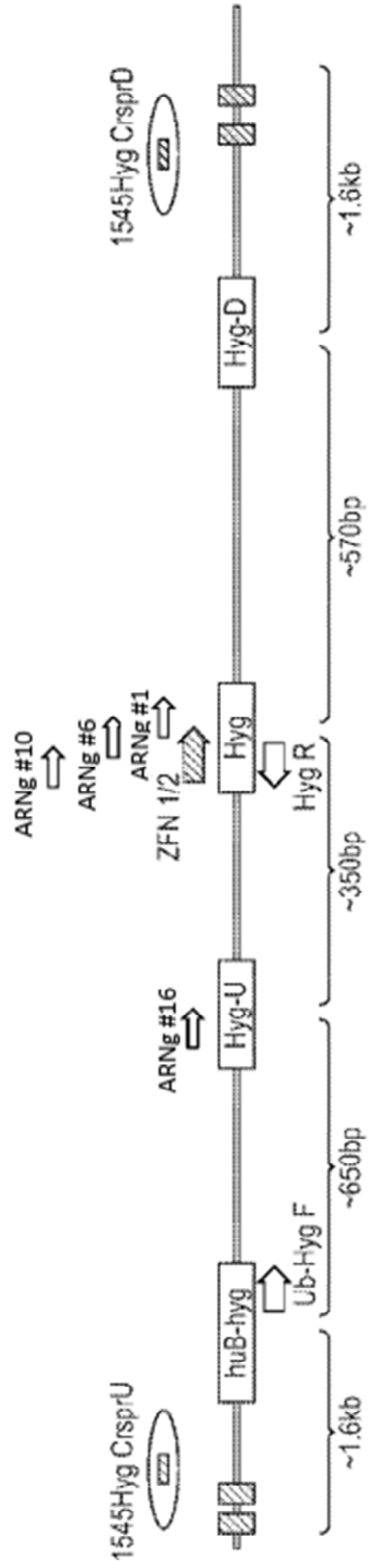


Figure 2



**Figure 3**

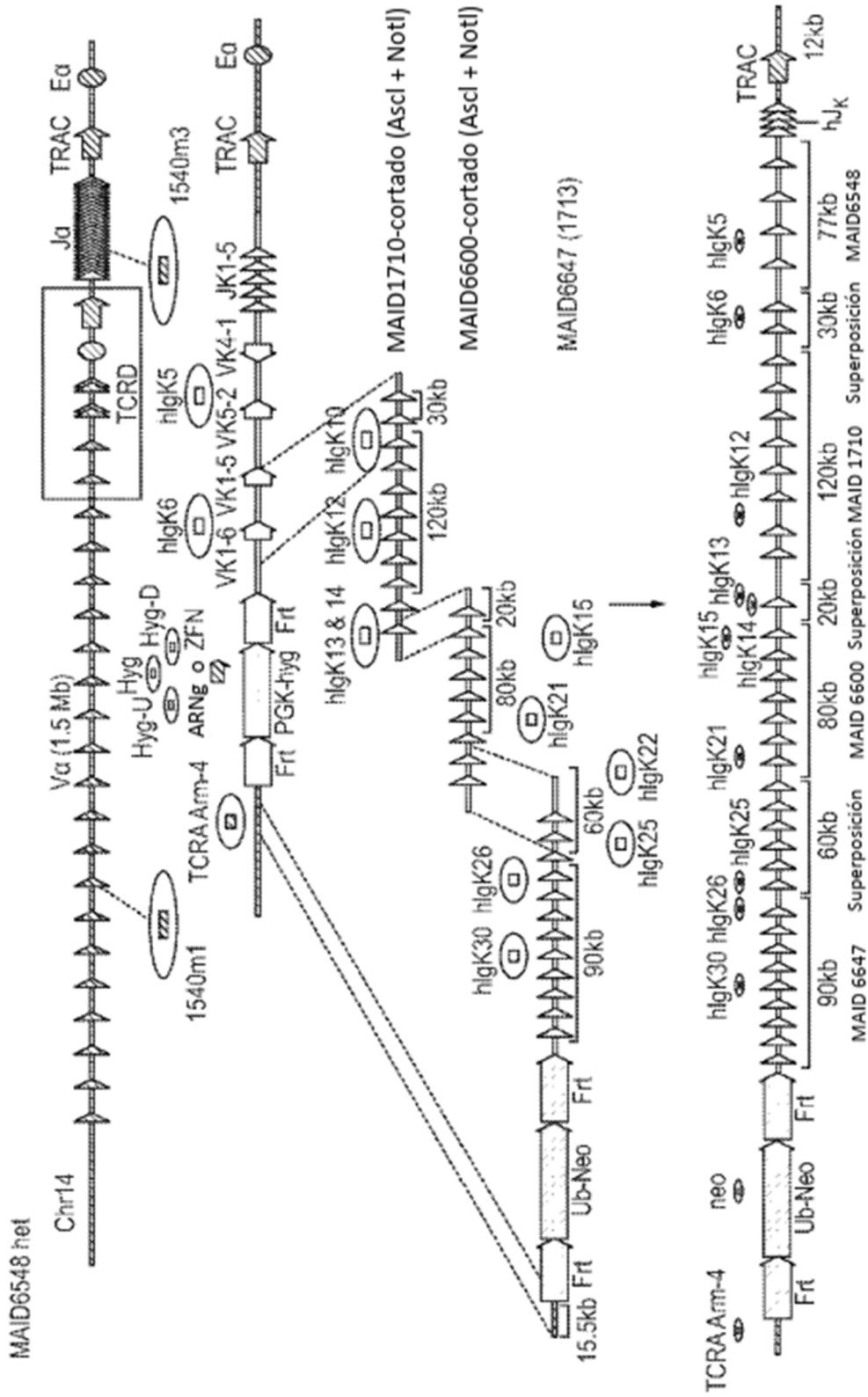


Figura 4